

## Joys of (top-notch) supervision

**The roles of mentors in research are seldom appreciated, let alone rewarded. All the more reason to celebrate the winners of the *Nature*/NESTA awards for creative mentoring in science.**

One hears horror stories about labs all too often. Graduate students who are left to sink or swim. Highly competitive environments in which people don't talk to each other. Labs where graduate students are at the beck and call of an overbearing lab head who treats them like fodder for his or her own glory...

Paradoxically or not, such tales often emerge from outstanding labs. It seems that 'incorrect' management and rank exploitation are no barrier to a record of outstanding scientific achievement — by the person doing the exploiting, at least.

Less often heard are tales of outstandingly good management. Most important, surely, are ways in which some lab heads empower their students not only to achieve outstanding work in their years of formation, but also to continue in that vein independently. Better still if the students hand on such empowerment to the next generation.

So we set out to look for such outstanding mentors.

We did so in collaboration with NESTA, the UK National Endowment for Science, Technology and the Arts. NESTA sits between the UK research councils and the arts funding agencies in supporting projects that foster individual artistic, technological and scientific creativity (see [www.nesta.org.uk](http://www.nesta.org.uk)). Its terms of reference required the awards to be restricted to Britain.

### Worthy winners

We decided to award two prizes. One, for a lifetime achievement in creative mentoring in science, goes to Tom Kibble, professor of physics at Imperial College London.

Whether through direct supervision or close collaboration at an early stage in their careers, the five distinguished researchers who nominated Kibble clearly felt a very close bond with him. He himself acknowledges the leadership of two other figures, Abdus Salaam and Paul Matthews, and says he learnt from them "the importance of creating not just a one-to-one relationship with a student but a group of students and postdocs who supported each other".

As his five nominating protégés testified, Kibble always made himself available, at least once a week. Some students are lucky if they see their supervisor more than once a year.

The key skill that students need to acquire, says Kibble, is the ability to select appropriate problems. His approach was to suggest initial problems to think about and then to encourage graduate students to find their own problems in their second or third years. His own work has been in the theory of the very early Universe and the development and observable consequences of defects in the ground-state structure of the vacuum.

The judges of the competition (see [www.nature.com/nature/nestaawards/index.html](http://www.nature.com/nature/nestaawards/index.html)) recognized some common characteristics of the shortlisted candidates. Many of these were well captured by one of Tom Kibble's nominators: "He has the knack of listening to others and giving his opinion so that one feels one's ideas are worthwhile. I have never heard of him being too busy to talk to someone. I have never heard him give the impression that someone's work was clearly wrong. Instead he would run it round and reformulate the idea so that it was something sensible to pursue." His protégés also appreciated the breadth and depth of his scientific insight.

Our other winner, of a mid-career award, is Innes Cuthill, professor of behavioural ecology at Bristol University, who designs innovative experiments to test questions about the selective forces that shape animal form and behaviour.

Again, protégés speak of an open door, and an ability to listen with a sharp mind and strike a balance between suggestions and leaving the student free to develop his or her own ideas — and also, an approach tailor-made for each student. As one of his nominators said: "He happens to be an incredibly original thinker and creative scientist who knows how to bring out these qualities in others. Mentoring isn't something Innes does in addition to being a scientist. It is totally integral to the way he does science."

"Innes understands that everybody works in their own idiosyncratic way," said another. "Throughout my career I have watched other colleagues being moulded to specific ways of working and thinking by their supervisors and lab leaders. But as far as I know, Innes has never done this with any of the countless undergraduate and postgraduate students who have passed through his door."

Mentors sometimes need imagination in finding the way to be supportive. When one PhD student asked for an assistant to help achieve a task, Cuthill said he would act as his research assistant: "Just tell me what to do." As Cuthill recalls: "He felt his supervisor had confidence, I was there on hand when he had doubts, and showed I wasn't expecting him to do anything I wasn't also doing myself. I have followed this model ever since."

### Top tips

*Nature* and NESTA hoped that the competition might expose not only the characteristics of mentors but also techniques that others might adopt. In the end, the message seems to be that personality plays the dominant role. But we encouraged entrants to suggest general principles, and it seems appropriate to record Cuthill's:

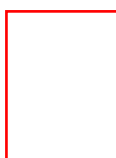
- Never let a student think they have asked a silly question. Otherwise you risk stifling enthusiasm and killing self-confidence.
- Get involved in the research. Be their research assistant. It shows your enthusiasm for their work more effectively than (just) checking stats and draft manuscripts. And you can spot technical errors and improvements that might otherwise be missed.
- Give students time to reach their own conclusions: do not immediately jump in with the solution.
- Remember to emphasize the positive. Young researchers don't know that draft manuscripts often need heavy reworking and that even good papers get rejected.
- Emphasize that science isn't personal. When someone criticizes your work, they are not criticizing you.
- A supervisor must earn co-authorship by doing more than the normal role of a supervisor.

*Nature* is delighted to congratulate these winners and, inspired by these and others nominated for the awards, will be highlighting the role of mentors in future issues. We expect to organize more mentorship awards, in Britain and elsewhere. In doing so, we seek to celebrate not only the best of science's laboratory culture, but also outstanding examples of its humanity. ■





**Peer review**  
Harvard president  
loses faculty  
confidence vote  
p424



**Brought to book**  
French digital library  
borrows from  
Google's plan  
p425



**Hard times**  
Academies' report  
shows a postdoc's  
lot is not a happy one  
p426



**Primate mover**  
Biologist becomes  
top brass at monkey  
centre  
p428

# US launches probe into sales of unapproved transgenic corn

**Colin Macilwain, Washington**

A strain of genetically modified corn that does not have regulatory approval has been distributed by accident over the past four years, *Nature* has learned.

Syngenta, one of the world's largest agricultural biotechnology companies, revealed the mistake to US regulators at the end of last year. Although the crop is believed to be safe, the fact that it was sold for years by accident raises serious questions about how carefully biotechnology firms are controlling their activities, critics say.

The corn (maize) was modified with a gene from the soil bacterium *Bacillus thuringiensis* (*Bt*), which is inserted into the crop to act as a pesticide. Syngenta has approval to sell a variety of the transgenic crop called *Bt11*, which has been used successfully for many years in the United States and elsewhere. The strain has been approved for consumption in the European Union, for example, and may be one of the first food crops approved for cultivation there.

But between 2001 and 2004, Syngenta inadvertently produced and distributed several hundred tonnes of *Bt10* corn — a different genetic modification that has not been approved.

Since the release was discovered in late 2004, US government scientists have assessed the *Bt10* corn — which differs from *Bt11* by only a handful of nucleotides on a section of the gene that does not code for the protein toxin — and have concluded that it is safe to eat and poses no environmental threat.

"What makes this somewhat unique is that *Bt10* and *Bt11* are physically identical and the proteins are identical," says Jeff Stein, head of regulatory affairs at Syngenta in Research Triangle Park, North Carolina.

Sarah Hull, a spokeswoman for the company in Washington DC, adds that Syngenta promptly reported the mistake to regulators after the discovery. She says this shows that the system is working as it should do. Company officials also note that the release was relatively small. About 150 square kilometres of the crop was planted over the four years,

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

**Seeds of doubt: critics say release of the wrong form of corn shows more regulation is needed.**

they say, which is 0.01% of all corn planted in the United States during that period. As *Bt* corn seed has to be bought every year, rather than being gathered from the previous year's crop, the problem should not escalate.

## Hard to swallow

But Michael Rodemeyer, director of the Pew Initiative on Food and Biotechnology, a think-tank in Washington DC, says that the release reflects the absence of a thorough monitoring system for genetically modified products in the US food supply. "This will raise questions in the minds of countries that import food from the United States about whether we have adequate controls in place," Rodemeyer says. "It will provide ammunition for critics of genetically modified food — and it may provide incentives for countries to look at non-genetically modified varieties."

Syngenta discovered the mistake when one of its seed manufacturers, which was attempting to use the corn seeds in plant-breeding experiments, informed it that the seed was not *Bt11*.

Syngenta then told the Environmental Protection Agency (EPA), the Food and Drug Administration and the US Department of Agriculture (USDA), which are jointly responsible for approving genetically modified crops. Regulators and the company have since been involved in months of discussions over what should be done about the error, and how and when information should be released to the public.

White House officials have also been involved in these sensitive talks, partly because the United States and the European Union are locked in a fierce trade dispute over whether tough European rules to trace the flow of genetically modified crops are scientifically necessary. Syngenta officials declined to list the countries that accidentally received the *Bt10* seed.

In a statement released to *Nature* on 14 March, the EPA says that regulatory agencies are "conducting investigations to determine the circumstances surrounding and extent of any violations of relevant laws and regulations". The EPA says that it is investigating whether the Federal Insecticide, Fungicide, and Rodenticide Act has been breached, and that the USDA is looking at possible violations of the Plant Protection Act. "The US government is also communicating with our major trading partners to ensure they understand there are no food safety or environmental concerns," it adds.

The last major, unintended release of a genetically modified crop in the United States occurred in 2000, when a *Bt* corn known as StarLink was inadvertently planted for human consumption. Because of possible allergic reactions, StarLink had been approved for use only in animal feed. Recall of StarLink corn cost the food industry an estimated US\$1 billion, according to Rodemeyer, and lent impetus to global concerns about the safety of genetically modified food.

# Obesity expert owns up to million-dollar crime

**Rex Dalton, San Diego**

A US physiologist is facing a possible prison sentence after pleading guilty to a felony in one of the largest research misconduct cases on record.

On 17 March, through a comprehensive agreement with the US government, Eric Poehlman acknowledged falsifying 17 grant applications to the National Institutes of Health (NIH) for nearly \$3 million, and fabricating data in ten published articles.

Poehlman, 49, who could not be reached for comment, has been barred for life from receiving US grants. He faces up to five years in prison and the ten articles are being retracted.

In January, Poehlman resigned his most recent position at the University of Montreal in Canada. He has published more than 200 articles on obesity, the menopause and ageing. His eight years of misconduct up to 2000 occurred when he was a researcher at the University of Maryland in Baltimore and

at the University of Vermont in Burlington, which initiated the investigation against him in 2001.

"This is the biggest case we have ever had," says Alan Price, head investigator for the Office of Research Integrity (ORI), which monitors grants for agencies such as the NIH. The only other criminal case of research misconduct in the United States involved psychologist Stephen Bruening of the University of Pittsburgh, Pennsylvania. In 1988 he received a 60-day sentence for providing false information about a drug study.

Authorities say Poehlman engaged in a variety of tactics to derail the investigation into his activities. All proved unsuccessful, and eventually led the US Attorney's Office in Vermont to seek the criminal conviction.

In 2001, Poehlman sued the university in federal court to prevent the ORI from being notified of possible misconduct. When he was unsuccessful, Poehlman quickly and

quietly departed for Montreal, where he was given a research position endowed by the Canadian government. Montreal was not aware of the allegations until the autumn of 2003, officials say, at which point he was placed on leave.

The first questions about Poehlman's research were raised at Vermont in late 2000 by a student researcher, Walter DeNino, who observed data being altered. Last year, DeNino filed a federal false-claims suit against Poehlman for misrepresentations made to the NIH. The US government then secured \$180,000 — virtually all of Poehlman's assets — as repayment for the falsifications to the NIH. DeNino is to receive \$16,000 to cover his legal fees.

"I was shocked when I discovered Dr Poehlman's activity," says DeNino, who is currently a student at Columbia University in New York. "He and his lawyers attacked my character." ■

# No-confidence vote fails to shift Harvard president

**Emily Singer, Boston**

Faculty members at Harvard University — encouraged by their own audacity in passing a vote of no confidence in their president — are planning to challenge him further.

Members of the university's Faculty of Arts and Sciences voted 218 to 185 against Larry Summers in a secret ballot on 15 March. Some say they will follow this by calling directly for his removal, and by censuring the Harvard Corporation, the executive board that appointed him.

Others are wondering how Summers will be able to function as president now

that the largest of Harvard's faculties has approved a motion stating: "The faculty lacks confidence in the leadership of Lawrence H. Summers."

"A weakened president will not be able to lead a fund-raising campaign," suggests anthropologist Arthur Kleinman.

An immediate change in Summers' status seems unlikely, however. The Harvard Corporation continues to support him, and Summers has given no indication that he will step down. In a brief statement made after the vote, he said he would continue to work with faculty to restore a sense of trust.

"Many of us hope a transformation will take place," says Everett Mendelsohn, a historian of science. He envisages a greater transparency in the dealings of the president — and more power for the faculty.

Lorand Matory, an anthropologist who sponsored the no-confidence motion, says he will call directly for Summers' removal at the next faculty meeting in April. There may also be a move to rebuke the corporation.

Summers came under fire after a speech on 14 January in which he suggested that differences in "intrinsic aptitude" between men and women might be a major factor behind the scarcity of senior female scientists.

Anger over the comments quickly blossomed into a broader attack on Summers' leadership style (see *Nature* 433, 790; 2005). Supporters and critics expressed surprise, however, at the outcome of last week's meeting, in which members also voted 253–137 for a milder rebuke.

Discontent on Harvard's campus in Cambridge, Massachusetts, is bringing other controversies to light. Last year, the Harvard School of Public Health won a five-year, \$107-million AIDS grant. According to a column in *The Boston Globe* last week, Summers and colleagues tried to take control of the project and to silence protests from its lead investigator. One senior faculty member — who declined to be identified — told *Nature* that the crisis over Summers' tenure had encouraged those involved to start talking. "Now they are less afraid," the professor says. ■



Summers' lease is all too short: Harvard's largest faculty has challenged the head of the university.



# France takes on Google in scanning race

**Declan Butler, Paris**

French president Jacques Chirac instructed his government last week to come up with proposals for digitizing the collections of libraries in France and other European countries.

His statement, issued on 16 March, asked Renaud Donnedieu de Vabres, France's minister of culture, and Jean-Noël Jeanneney, the president of the Bibliothèque nationale de France, to come up with proposals to accelerate the dissemination of French and other European works on the Internet. He called on France and Europe to take "a major role" in a "vast digitization of knowledge".

Chirac's move is widely interpreted as a response to Google's announcement late last year that it intends to scan millions of library books — primarily from collections at the universities of Harvard, Stanford, Michigan and Oxford, as well as that of the New York Public Library — over the next ten years. But this plan is being viewed with trepidation by backers of existing, public-domain projects that aim to do the same kind of thing.

One backer of the public-domain approach is Brewster Kahle, founder of the Internet Archive project, based in San Francisco. In December, Internet Archive teamed up with Carnegie Mellon University, the Library of Congress American Memory Project and universities in Canada, Egypt, India, China and Europe to digitize 9 million books over the next four years. More than 50,000 of them will be digitized by the end of this month.

Kahle says that the Google project could have three possible outcomes. The first is that

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

**Text appeal: Jacques Chirac urges the public digitizing of books in response to Google project.**

funding for public-domain projects could dry up, with library collections effectively being privatized by Google. Alternatively, the Google move might result in healthy competition and an increased demand for a public-domain service, Kahle says. He cites as a precedent the human genome project, where the private company Celera's plans to sequence the genome galvanized the public consortium's determination to deliver its own version. The third possibility, Kahle says, is that Google might collaborate successfully with the public-domain efforts.

The Internet Archive's annual administrative costs of about \$2 million are met by grants from the US National Science Foun-

dation, the Library of Congress, national archives such as those in Britain and France, and philanthropists such as the William and Flora Hewlett Foundation. But its scanning costs — which could amount to \$230 million over four years — are due to be paid by participating libraries. There is now "fear, uncertainty and doubt" over this, says Kahle, with some libraries "waiting to see if they can get a handout from Google" instead.

Michael Hart, the founder of Project Gutenberg — the first ambitious attempt to digitize libraries, launched in 1971 and based in Urbana, Illinois — expresses concern about the proprietary nature of the Google project. He fears that his and other public projects could be hurt if funders think Google can do the job alone.

A public effort is essential, argues Hart, because it should provide users with access to the full text of books and high-quality images that they can use in whatever way they wish, without restriction. In contrast, Google's current system allows users to search texts online and to browse images, but provides access to only a small portion of the texts.

However, Raj Reddy, a computer scientist who is the founder and director of the Universal Digital Library at Carnegie Mellon University, welcomes the competition from Google, and says that, if anything, it should increase support for public-domain projects. "The more the merrier," says Reddy, whose project scans a million pages a day and has already indexed some 100,000 volumes as part of a plan to scan one million books. "It is going to take us a long time to digitize all these things." ■

## Surfeit of boys could spread AIDS in China's cities

**David Cyranoski, Beijing**

China's AIDS crisis could take a turn for the worse because of a surplus of men in the population, suggests a new study.

A paper published in this week's issue of the journal *AIDS* (19, 529–547; 2005) argues that an excess of boys in China's population could exacerbate the AIDS problem in cities. The sex bias is the result of the nation's 'one-child' policy and the fact that many parents want that child to be male.

The paper, "Surplus men, sex work, and the spread of HIV in China", is authored by Joseph Tucker, a medical doctor in training at the University of North Carolina's Center for Infectious Diseases, and his co-workers. It says that there are now about 12 boys born for every 10 girls in China. The widespread availability of ultrasound techniques, which allow doctors to identify the sex of fetuses

and give parents the ability to choose, is worsening the imbalance. Tucker argues that the resulting "surplus men" — which he calculates to be about 8.5 million — are likely to be unmarried, poor and lacking education.

The number of prostitutes in China has also skyrocketed: it has admitted that there are now between 4 million and 6 million, compared with only 25,000 in 1985.

Historical precedents, such as a sex-ratio imbalance caused by mass migration to Shanghai in the 1930s, have led to rampant sexually transmitted infections. Tucker, who spent a year working at a clinic for sexually transmitted diseases in Nanjing, says that another such situation could be developing.

Until now, consideration of the HIV problem in China has often focused on isolated communities, such as intravenous drug users or the victims of contaminated

blood banks. But sexual transmission is playing an increasing role and some fear it will lead to an urban explosion of AIDS.

Some epidemiologists and HIV experts have challenged the speculative nature of Tucker's argument, however. Terry Hull, a social demographer at the Australian National University, Canberra, questions the idea that "poor unemployed migrant males" will be fertile ground for HIV transmission. "We might as easily argue that they will also ignore safety messages on the job and have more occupational deaths," he says.

Tucker admits that his theory is only a hypothesis. But he hopes it will focus more attention on a possible seed of disaster. He calls for more education at construction sites, military areas and unemployment centres, where the surplus males gather. "These boys are coming of age," he says. ■

# Academies seek better prospects for postdocs

Jessica Ebert, Washington

The US National Institutes of Health (NIH) is being urged to introduce a set of reforms to improve the lot of postdocs and increase their chances of establishing independent scientific careers.

A report from a National Academies panel chaired by Thomas Cech, president of the Howard Hughes Medical Institute in Chevy Chase, Maryland, says that postdoctoral training should be limited to five years.

The 18 March report also says that postdoc training grants should be made available to non-US citizens, and that money should be transferred from the NIH's main funding mechanism to support independent research awards that would enable postdocs to pursue their own projects.

Cech says that, if implemented, the recommendations will "encourage young investigators to take on top-rate, important projects, to deviate from the research of their previous mentors, and to strike out in new directions, tackle new systems and develop new methods".

The study also suggests that the NIH should introduce a class of investigator grant specifically tailored to help young scientists moving into their first posts as independent investigators. It recommends awarding 200 of them annually, worth \$500,000 each over five years.

"The current NIH grant system really honours safe research in well-established pathways," Cech claims. "We want to break away from that and find mechanisms that will free up researchers in the early stages of



Rocky road: many US postdocs find that their chosen career path can be tough going.

their career to make the big discoveries that are really going to have an impact on medicine and on human health."

Last June, NIH director Elias Zerhouni asked the National Academies to look into mechanisms that would enhance postdoctoral training and foster young researchers' independence.

Zerhouni says that some of the committee's recommendations — such as that requiring senior researchers to describe in their grant applications what they would do to nurture their postdocs' careers — could be implemented relatively quickly. He notes that a steering committee to deal

with the other recommendations has already been set up.

The report was requested after it emerged that the NIH budget is increasingly going to fund older, established researchers. Between 1980 and 2003, for example, the percentage of grants going to researchers aged under 40 fell from more than a half to less than one-fifth.

"We've come up with recommendations that are evolutionary, not revolutionary," says Cech, who hopes this will make it easier for the NIH to implement them. "The vitality of the biomedical sciences in the United States depends on taking some action soon," he adds. ■

## US undervalues foreign researchers, survey reveals

Rex Dalton, San Diego

Foreign postdocs in the United States work longer hours and publish more, but are paid less than their American counterparts, according to a poll of young scientists.

The survey, by Sigma Xi, a scientific research society in Research Triangle Park, North Carolina, also found that substantial numbers of foreign postdocs experience some form of visa or international travel problems because of US security rules.

But a preliminary version of the analysis, which was released on 11 March at the third annual meeting of the National Postdoctoral Association in San Diego, did contain some encouraging results. Male and female postdocs are paid about the same, it indicated, and stipends have risen steadily over the past decade. "That is pretty good news," says mathematician Geoff Davis, who led the survey.

The questionnaire gauged the views of 7,500 postdocs at nearly 50 universities or research institutions. It was sent to 22,000 postdocs — about 40% of the total thought to be working in the United States.

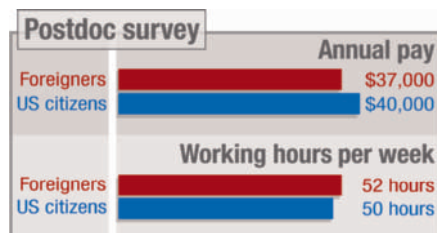
The median salary for all respondents was \$38,000 per year — up from \$28,000 in 1995. On average, electrical engineers fared best, earning about \$45,000, and ecologists came off worst with \$35,600.

The median for foreign postdocs was \$37,000 — 8% less than the \$40,000 received

by US citizens. But the foreign scientists said that they worked 52 hours a week — two hours more than the Americans. And international postdocs said that they had produced almost 30% more peer-reviewed, published articles than their American counterparts.

Nearly three-fifths of the foreign researchers reported experiencing difficulties on re-entering the United States after travelling abroad.

The survey, which is published in full in *American Scientist* next month, comes at a time when US universities and research agencies are making increasing efforts to address the lot of postdocs (see above). Faced with postdoc organizations springing up on campus, many universities are appointing administrators to operate offices dedicated to them. But postdocs continue to complain of long hours and uncertain career prospects. ■



# Lawsuits and logistics tie up California's stem-cell funds

Peter Aldhous, San Francisco

Biologists anticipating a stem-cell gold rush in California will have to wait a little longer, as a string of setbacks is slowing the state's plans to jump-start such research.

California's voters agreed last November to bankroll a \$3-billion, decade-long initiative for research on stem cells, and backers had hoped that the first research grants would appear by May of this year. But the initiative has come under mounting political and legal fire. Its leaders, facing the daunting task of building a substantial research agency from scratch, now concede that grants won't begin to go out until autumn, at the earliest.

The stem-cell ballot measure, known as Proposition 71, was drafted by Bob Klein, a property developer whose son suffers from juvenile diabetes. Klein is determined to get money into scientists' hands as quickly as possible. "We need to do our best with a responsible but aggressive schedule," he told *Nature* in late February.

To minimize red tape, Proposition 71 made supervision of the new agency, the California Institute for Regenerative Medicine (CIRM), the responsibility of a 29-person citizens' committee. Klein was quickly installed as committee chairman. But critics have attacked what they say is a lack of transparency in the committee's procedures, and have raised concerns about conflicts of interest. Committee members include major players in the biotechnology industry and leaders of academic institutions that are likely to apply for CIRM funding.

On 14 March, Bill Lockyer, the state attorney-general, acknowledged in a legal brief that the sale of bonds to finance the CIRM cannot begin until two lawsuits challenging its constitutional legality have been resolved. The suits have been brought by taxpayer groups and opponents of stem-cell research, and will put any grant-making on hold.

Two days later, state legislators announced moves to open up the CIRM to greater public scrutiny, and to restrict donations of human eggs for research. State senators Deborah Ortiz, a Democrat, and George Runner, a Republican, are proposing a constitutional amendment that would tighten the CIRM's conflict-of-interest rules, and force Klein's committee to hold all of its meetings in public. To pass, the amendment requires a two-thirds majority in both houses of the California legislature, and must then win support in a public ballot.

Ortiz and Runner are also proposing a three-year moratorium on women donating multiple eggs for research. The supply of human eggs is the chief limiting factor

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

Tricky proposition: Bob Klein's plans to see research grants issued by May will have to wait.

in cloning research. "It'll be a problem if there are no egg donors," warns Lawrence Goldstein, who studies neurodegenerative diseases at the University of California, San Diego.

The proposed legislation could still allow some eggs to be donated by women who are undergoing *in vitro* fertilization. If so, says Evan Snyder, a stem-cell researcher at the Burnham Institute in La Jolla, it should be possible to begin cloning studies. But as *Nature* went to press, the precise wording of the bill had not been released.

CIRM officials say that Ortiz and Runner's planned constitutional amendment will not affect the awarding of grants. But the lawsuits could cause delay. The suits contend that the CIRM is unconstitutional because it sidesteps state laws on governance and conflicts of interest. Lockyer has asked the Supreme Court of California to deal with the suits quickly, but it is unclear when they will be resolved.

Even before these developments, the practical challenges of establishing a new granting agency from the ground up had forced the CIRM to retreat from the idea of issuing grants in May. The first grants would go to institutions to allow them to hire graduate students and postdoctoral researchers. Zach Hall, an experienced science administrator who is the CIRM's interim president, told reporters on 15 March that such grants are unlikely to be approved until after the summer, irrespective of developments in court. ■

## Fake papers hamper plans for nuclear store at Yucca Mountain

Geoff Brumfiel, Washington

Documents relating to the safety of the proposed nuclear-waste repository at Yucca Mountain in Nevada may have been falsified, the US Department of Energy (DOE) revealed last week.

According to agency officials, a hydrologist at the US Geological Survey (USGS), who was studying how water flowed through the mountain, faked documentation on the times and dates at which certain geological samples were taken from the site.

The independent inspectors-general at both the USGS and the DOE have launched investigations, and the Nevada congressional delegation has called for a federal investigation into the issue.

The Yucca plan would see canisters of nuclear waste stored under the mountain for thousands of years. The water data are crucial because the repository's safety depends heavily on how wet it is likely to get inside, and on how long it takes for water to escape from the mountain rock into surrounding river systems.

A 1996 study by Los Alamos National Laboratory in New Mexico indicated that water takes just a few decades to flow through the mountain, raising fears that storage containers might erode and leak nuclear waste into the water table. But the DOE argues that subsequent studies and models by USGS scientists show that water moves more slowly through the mountain and that there is little threat to the local water supply.

Evidence for the falsification surfaced while the DOE was drawing up an application for a licence to store nuclear waste at the site. E-mails between the hydrologist and several colleagues dating from 1998–2000 were found that indicated he had faked documentation on the data and some computer models used to predict water flow through the mountain, says a USGS spokeswoman.

"What it means scientifically we don't yet know," says John Garrick, chairman of the Nuclear Waste Technical Review Board, which oversees the Yucca project. "But there is no doubt that it's affected public confidence in the project."

Senator Harry Reid (Democrat, Nevada), the minority leader in the Senate who opposes the construction of the repository, is calling for a full federal investigation into the alleged misdeeds. "It is abundantly clear that there is no such thing as 'sound science' at Yucca Mountain," he says. ■



## Congress threatens ITER over cost to US fusion labs

**Washington** History is threatening to repeat itself as members of the US Congress chafe at the country's participation in an international fusion experiment.

The United States pulled out of the ITER project in 1998 after Congress expressed frustration over its high price. George W. Bush's administration opted to re-enter the collaboration in 2002 (see *Nature* 421, 563; 2002). But at public hearings on 15 March, members of Congress expressed concern about the project, which has been deadlocked over the choice of site since 2003.

Bush's 2006 budget proposal allots \$50 million to ITER, but reduces domestic fusion funding by some \$34 million. "I feel that everyone is bowing to ITER," says Rodney Frelinghuysen, Republican representative for New Jersey. "It doesn't make me happy."

Senator Pete Domenici (Republican, New Mexico) also has reservations about the project. The appropriations subcommittee has been "very clear in its opposition to funding ITER out of the fusion account", he said in a statement last week. Stephen Dean, who heads a fusion lobby group in Washington, says that funding ITER at such expense to domestic programmes "is just not acceptable to the fusion community".

## Farm-scale trials reveal effects of transgenic crops

**London** The final results from the largest investigation so far into the ecological impact of genetically modified crops have revealed a subtle impact on biodiversity.

Data from the four-year Farm Scale Evaluations, commissioned by the UK government and carried out in Britain, show that the herbicide spraying regime associated with transgenic winter oilseed rape favours grassy weeds at the expense of broad-leaf rivals.

Researchers led by David Bohan of Rothamsted Research in Harpenden reported on 21 March that the decrease in broad-leaf weeds might adversely affect

## Change of head lifts gloom at primate centre

**Durham** Researchers at the Duke University Primate Center in Durham, North Carolina, are hoping that the appointment of a director will give the centre a new lease of life. Anne Yoder, a biologist at Yale University who got her PhD at Duke University, will take over the running of the centre in August.

The centre, which combines research and conservation missions, is well known for its collection of endangered prosimians, such as lemurs. But two years ago its future was threatened when university administrators considered withdrawing their commitment to the facility (see *Nature* 423, 471; 2003).



The 16 March announcement to hire Yoder — along with a commitment to spend \$7 million expanding and restructuring the centre to capitalize on evolutionary genomic opportunities — seems to signal a brighter future for the facility.

insects and birds that rely on such weeds and their seeds. But the overall effect on wildlife is less pronounced than that previously indicated for spring oilseed rape and beet (see *Nature* 425, 751; 2003).

## Panel holds the key to prison research rules

**Washington** The Institute of Medicine (IOM), which advises the US government on biomedical issues, is casting a critical eye on the rules controlling research on prisoners.

The regulations were put in place by the Department of Health and Human Services (DHHS) in 1976, and define tricky concepts such as informed and voluntary consent. But researchers now say the rules are difficult to follow and implement. In addition, the number of inmates in jails and juvenile detention facilities has risen from 200,000 to more than 2 million since the 1970s.

The DHHS has now asked the institute to reconsider the ethical bases for protecting prisoners involved in research. An IOM committee, led by public-health lawyer Lawrence Gostin of Georgetown University in Washington, will address such questions as how the term 'prisoner' should be defined, whether it is possible to ensure informed and voluntary consent in prisons, and whether research on prisoners should only be done if it benefits prisoners — as is currently the case.

## Environmentalists aim to foil Swiss plan to wrap glaciers

**Munich** When the artists Christo and Jeanne-Claude wrapped the Berlin Reichstag in fabric ten years ago, the project catapulted them to stardom. A Swiss company has now said it wants to wrap an alpine glacier in PVC foil this summer. But far from being seen as high art, the plan — intended to stop

the glacier from melting — has triggered protests from environmentalists.

Climate warming is causing the Gurschen glacier, near the Swiss ski resort of Andermatt, to shrink. In a bid to slow the melting, researchers will in May cover part of the glacier with a material that reflects sunlight. If the test is successful, the use of the foil covering may be extended next year, says Carlo Danioth, head of piste services at the Andermatt funicular.

But Swiss environmentalists complain that the project will be an eyesore. Wrapping the glacier to stop it melting is an "absurd idea", says a climate expert at Greenpeace in Switzerland.

## Congress to consider ban on human cloning

**Washington** In what has become an annual tradition in the US Senate, lawmakers have introduced legislation to ban human cloning for both reproductive and research purposes.

On 17 March, Sam Brownback (Republican, Kansas) in the Senate and Dave Weldon (Republican, Florida) in the House of Representatives introduced bills banning all forms of cloning. The bills would have to be passed by both houses of Congress, and be signed by President George Bush, before becoming law.

Members of the House who support therapeutic cloning — in which the cloned embryo is used to make cells either for research or transplantation — have already introduced a bill that would permit this but ban reproductive cloning. Senators are now expected to do the same, although in previous years neither side has garnered enough votes to pass legislation on therapeutic cloning. Brownback's supporters hope that Republican gains in the November 2004 elections will now tip the vote in his favour.

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

On trial: studies of oilseed rape have shown that transgenic crops can affect biodiversity.



**B**irthday parties don't get much cooler than this — a mountain-top get-together for researchers fascinated by ultracold matter. Last month, to mark the tenth anniversary of the creation of the first Bose–Einstein condensate, physicists met in Banff, Canada, to discuss ultracold atoms in the morning and ski in the afternoon. The air outside was cold and crisp, but inside the atmosphere was invitingly warm. The scene verged on the cosy. A Nobel prizewinner could be found sitting next to a gaggle of graduate students, and speakers were interrupted by good-natured questions. Indeed, Kathy Levin, a theorist from the University of Chicago, began her talk by saying: “There is a wonderful *esprit de corps* and camaraderie here — which one doesn't see in all fields, and which I think is a secret of its success.”

Levin recently escaped from the notoriously combative field of high-temperature superconductivity, another branch of condensed-matter physics. But her transition is not unique. In the past two years, with research into high-temperature superconductors stalled, more and more condensed-matter physicists have begun studying cold atoms. One is Fei Zhou of the University of British Columbia, who chaired a session in Banff and says he has never looked back since making the switch 18 months ago. Until then, Zhou says that every year he'd find fewer colleagues at the condensed-matter symposia he attended. By contrast, the field of ultracold atoms has swelled to some 100 labs and counting. There is excitement, funding and, most importantly, the chance to do some fascinating new science.

The concept behind Bose–Einstein condensation is much older than this youthful exuberance would suggest. In 1924, Albert Einstein and Satyendra Nath Bose used quantum mechanics to describe what would happen to a cloud of gas atoms if they were made so cold they essentially stopped moving. Squeeze them and they would merge into a single entity, a giant superatom. Locked together, moving as one, this condensate of atoms would become a new phase of matter — different from solid, liquid or gas.

This idea remained no more than a thought experiment until 1938, when helium-4 was cooled to below 2.2 K and became a new kind of fluid that flows without friction — a ‘superfluid’. But supercooled

helium is a liquid rather than a gas, and so is not considered a ‘true’ Bose–Einstein condensate (BEC). Back then, creating the much colder temperatures necessary to make a gaseous condensate seemed impossible.

Then, in 1995, two groups did it almost simultaneously. Eric Cornell of the National Institute of Standards and Technology (NIST) and Carl Wieman of the University of Colorado in Boulder cooled 2,000 rubidium atoms into one entity; and Wolfgang Ketterle, a physicist at the Massachusetts Institute of Technology, made a condensate from half-a-million sodium atoms. These feats were recognized with a physics Nobel prize in 2001, but no one had foreseen just how much they would inspire a new generation of physicists.

“The past ten years have just been an explosion,” says Ketterle, who was unable to make it to Banff. “There have been so many surprises. When we discovered the BEC we had a short list of what we thought would be important. What has been done by far exceeds our expectations — in even my boldest dreams I could

not think of so many interesting studies.”

Ketterle's surprise is understandable. Early research focused on making BECs — always a painstaking exercise — from yet more, and different, atoms. Creation was an end in itself. But over the past five years, BEC physics has grown, not just in size but in ambition. Today, although some researchers continue to characterize BECs, others attempt to apply BEC physics to other fields, and yet others are exploring the relatively new area of condensates made from a class of fundamental particle known as fermions.

An early question was whether BECs, like helium, are superfluids. In theory, once a superfluid starts swirling it should continue forever. Cornell and Wieman<sup>1</sup> first created such everlasting vortices in 1999, and today

most people accept that BECs are superfluids. At Harvard University, physicist Lene Hau<sup>2</sup> recently formed both vortices and ‘straight density waves’, which are akin to a sound wave, in the same BEC. She watched them collide and blossom into something like a spinning

**“What has been done by far exceeded our expectations — in even my boldest dreams I could not think of so many interesting studies.”**  
— Wolfgang Ketterle

# Some like it cold

In 1995, scientists created the first ultracold quantum gas and to their surprise launched a new scientific field. Ten years on and its chilly revelations are attracting a growing number of physicists. Karen Fox joins the party.

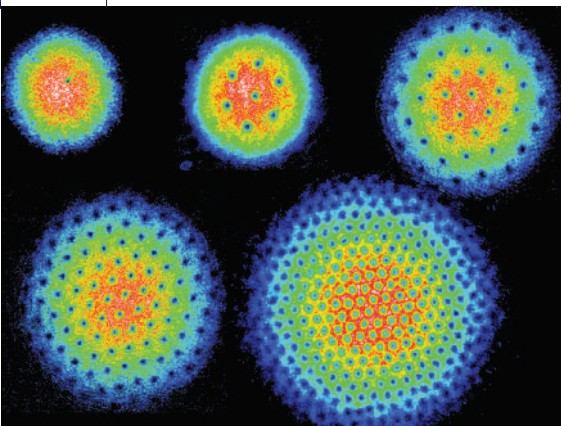


D. LEE

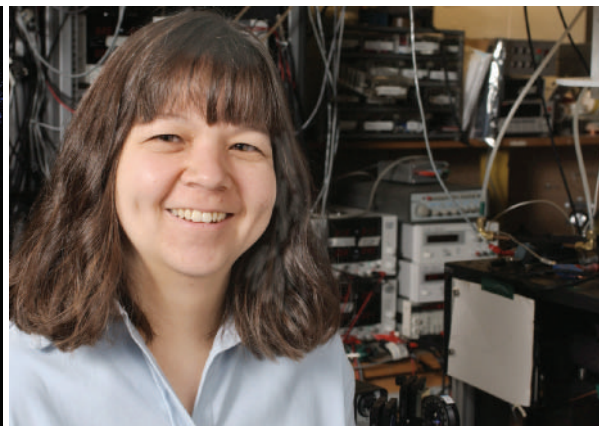


Physicists met in Banff (left) for the birthday of the first BECs, created by (left to right) Carl Wieman, Wolfgang Ketterle and Eric Cornell.

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS



Young physicists like Debbie Jin drive a fast-moving field that has seen breakthroughs such as BEC vortex lattices (left).



umbrella in the process of turning inside-out.

Such exquisite experiments demonstrate the control researchers now have over BECs. They have learned how to fine-tune the attraction between atoms in a BEC to do anything from spacing them in ordered lattices to forcing an entire BEC into something like a mini supernova. "I'm just amazed," says Ketterle, "that an experiment I thought was bloody difficult when I did it ten years ago is now being done regularly, with sophistication and much more experimental control."

And although theorists have had ten years to catch up, experiments are still driving the field. The moment a new theory comes out, there is someone in the wings ready to test it. "The experimentalists are wonderful," says Gordon Baym of the University of Illinois in Urbana-Champaign. "Not only do they do great experiments, but they do them once a week." It is this speed that helps make the field such a welcoming place. There is a sense that there is room for everyone.

Some of the fastest-moving research in cold atoms involves fermions. At a fundamental level, all matter comes in two varieties: fermions, such as protons, and bosons, such as photons. Creating condensates out of fermions instead of bosons was a highlight of the past two years. A star of fermion physics, Deborah Jin, a physicist with NIST in Boulder, discussed her latest work at the Banff meeting. As she describes it, one of the main differences between fermionic and bosonic matter is that two bosons can exist in the same place, like two crossing beams of light, whereas fermions cannot. This difference should rule out a fermion condensate.

But fermions can also behave strangely at low temperatures. When certain solids are cooled sufficiently, their conducting electrons — which are fermions — can pair up to create a boson, a phenomenon at the heart of our understanding of superconductivity. According to the 1957 Bardeen-Cooper-Schrieffer (BCS) theory, when two electrons merge to form a 'Cooper pair', this new boson can then effortlessly speed through the superconductor without resistance.

Early last year, Jin's lab reported a breakthrough — they had created bosons out of fermions and then turned them into a condensate<sup>3</sup>. From the start there were hopes that studying fermion condensates would give insight into BCS theory. This theory explains superconductivity only at very cold temperatures — high-temperature systems discovered in the 1980s spawned numerous theories and rancorous debate, but no firm explanations. It is perhaps this, more than anything, that has attracted so many condensed-matter physicists to the BEC field. As Randy Hulet, a physicist at Rice University in Houston, Texas, says: "We can use BECs to model condensed-matter systems so cleanly, in ways you just can't do in real condensed-matter systems."

Jin now hopes to 'see' a Cooper pair in a BEC — something researchers can only do indirectly in a superconductor. So far, the enigmatic dance partners remain hidden, but she can see correlations in the fermions'

positions that hint at Cooper pairs. Although this work isn't conclusive, it offers hope that studying the links between BEC and BCS behaviour will deliver new insight. "BCS theory is just very robust within condensed matter," says Levin. "It is the theory of all theories — if it turns out to be part of an even bigger theory, which I believe it is, then that's very exciting for us."

Hulet, a veteran of cold-atom research, is also interested in how BECs can influence other fields. He was one of the first to create a BEC soliton, essentially a wave that never dissipates. These intense waves might one day be used to make inertial sensors for detecting changes in gravity or acceleration. Today's inertial sensing relies on optical instruments and Hulet believes that cold atoms could increase their accuracy. But he admits, "So far the solitons we can produce are too small."

The greater reliability of atomic systems compared with optical ones has even got the military interested. Major Jay Lowell at the US Defense Advanced Research Projects Agency (DARPA), which funds cold-atom research, says he would like to see inertial

navigation systems based on BECs within five to ten years. DARPA's plans involve atom interferometer technology, pioneered by Ketterle's lab among others. An atom interferometer requires two coherent atom waves (essentially two atom lasers), which can then be made to overlap and produce an interference pattern.

DARPA is also tracking the potential of BECs in quantum computing — although Lowell's hopes for that are more in the 10–50 year range. Ignacio Cirac of the

Max Planck Institute of Quantum Optics in Garching, Germany, gave a fascinating talk in Banff on how one might get BECs to behave as qubits — the basic building block of a quantum computer. For the moment, however, the idea remains firmly in the realms of theory.

Predicting which directions will bear more fruit — fermions, solitons or something unknown — is not easy, but a field that even insiders consider esoteric continues to generate excitement. Young physicists who were at graduate school when the first BECs were made are now claiming the field as their own, and many older scientists are switching to what they hope are greener pastures. Ten years on, there's no slowing down. "I thank my lucky stars," says Hulet, "that I stumbled on to this intellectual gold mine."

**Karen Fox is a freelance writer based in Washington DC.**

1. Matthews, M. R. *et al.* *Phys. Rev. Lett.* **83**, 2498–2501 (1999).
2. Ginsberg, N. S., Brand, J. & Hau, L. V. *Phys. Rev. Lett.* **94**, 040403 (2005).
3. Regal, C. A., Greiner, M. & Jin, D. S. *Phys. Rev. Lett.* **92**, 040403 (2004).



# Looking for the ancestors

The scientists who discovered a new species of human in Indonesia last year are now back, looking for the bones that will flesh out their theories. Rex Dalton joins them.

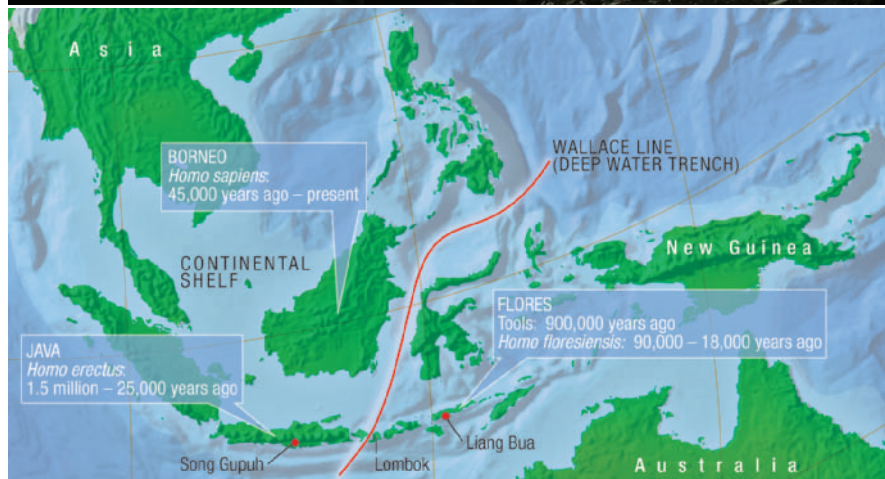
In a limestone cave on Java, sheltered from the monsoon rain that patters on the jungle leaves outside, a neatly dug shaft drops into the floor. Descending a bamboo ladder, workers disappear into the shadows to reinforce the pit's wooden supports. The archaeologists are anxious to begin extracting earth and, they hope, bones of hominins — species from the same family as modern humans. But first they have to pull out a huge rock that blocks their progress some five metres down — probably part of the ceiling that collapsed maybe 50,000 years ago. Thomas Sutikna surveys the situation with his colleagues from the Indonesian Centre for Archaeology. Gazing into the shaft, team co-leader Michael Morwood of the University of New England, Australia, dreams of the treasures that may lie beneath. "I think we have some very old material under this," he says.

Last year, this same group of Indonesian and Australian archaeologists electrified the world with their report of the discovery of a miniature human species, *Homo floresiensis*, which lived on the nearby island of Flores at least as recently as 18,000 years ago<sup>1</sup>. This new hominin, whose near-complete specimen was called LB1 and nicknamed 'hobbit' by some of the finders, represents a totally unexpected branch of the human evolutionary tree. The team is now hoping to unearth a similarly startling new find — the ancestors of *H. floresiensis*.

But a tense political backdrop may make the fieldwork difficult. At least one prominent Indonesian scientist publicly disputes the team's original description of the find; against prevailing opinion he says that the specimens do not represent a new species. And the discovery team claims that after



Digging deep: archaeologists search for traces of *Homo floresiensis*' ancestors in Song Gupuh cave.



he took the bones for study, they were returned in a damaged state that may hinder further analysis.

All this is enough to make the team seek a place of solace. So it is perhaps fitting that the researchers are currently digging in a cave called Song Gupuh (pronounced song-ga-poo) — 'flee cave' in English.

Here, in a cave where generations of people have sought refuge, the team is trying to

put controversy aside to concentrate on the next chapter of its research. The goal is to identify a transitional species — the hominins who later evolved into the hobbit's clan.

This should help anthropologists understand from whence *H. floresiensis* came, and perhaps how it acquired such a baffling mix of features: a metre-tall stature, ape-like arms and a small brain, but sufficient intelligence, seemingly, to make and use stone tools.



The cave is promising: it is very habitable, with water nearby and a good view of incoming predators. And it holds a deep accumulation of sediments, which may extend back some 100,000 years.

In some ways Song Gupuh is similar to Liang Bua, or 'cool cave' — the shelter where *H. floresiensis* was found nearly six metres under the floor. Both caves were partially excavated by an earlier generation of Indonesian researchers, before Morwood and his colleagues brought their teams to dig more extensively.

### Crossing the line

But Song Gupuh is also markedly different. Java lies on the western side of a 350-metre-deep undersea canyon, which played a crucial role in humanity's trek out of Asia. During glacial periods, tens or hundreds of thousands of years ago, ice covered much of the upper globe; the ocean level was so low that early humans could walk to Indonesian islands such as Java on the exposed shelf (see map). But Flores, along with other eastern islands, was isolated by deep water. Early humans could only reach these islands using some form of watercraft.

This isolation may help to explain why the features of *H. floresiensis* are so distinct. The species is thought to have grown small because of a lack of resources on the island. Fossils in the Liang Bua cave show that other animals were subjected to similar pressures: alongside the hominins, the remains of pygmy elephants were found.

However, the watery trench — named the Wallace Line after Alfred Russel Wallace, a contemporary of Charles Darwin who first noted the marked differences between animals living either side of it — also poses a problem. No one knows how hominins in early periods crossed the open water. During a ferry ride across the choppy waters of the Wallace Line to Lombok, Morwood marvels at the feat. "Do you think you could make it across here in a raft?" he asks.

It was this conundrum that first brought Morwood and his colleagues to the area. In 1996, Morwood's team reported stone tools at a site on Flores called Mata Menge. The site was dated to 900,000 years ago<sup>2</sup>, which is long before the period when anthropologists thought that sophisticated hominins lived there.

Morwood established a collaboration with researchers from the Indonesian Centre in Jakarta, including Sutikna. By looking for fossils on both sides of the Wallace Line, Morwood and his colleagues hoped to learn who made it across the water, when they got across, and possibly even how. One good candidate was *Homo erectus*, who is thought to have survived on Java until as recently as 25,000–50,000 years ago<sup>3</sup>.

By 2001, Morwood and his Indonesian colleagues were returning to caves that had been explored in previous decades by Radien Soejono, the senior archaeologist at the Indonesian Centre. Now 78, Soejono has ceded fieldwork to younger Indonesian colleagues. But he remains a national political and scientific force, and he was made co-chief investigator of the team with Morwood.

During the summer dig season in 2003, the researchers hit the big time at Liang Bua when they uncovered the LB1 specimen. Their original question about who made the 900,000-year-old tools paled in comparison. Everyone on the team was ecstatic — but their work set in motion rivalries in Indonesia that continue today.

Within days of the Liang Bua specimen's extraction, Soejono sought to have the bones transferred to the laboratory of his long-time colleague Teuku Jacob, a palaeoanthropologist at Gajah Mada University in the old Javanese capital of Yogyakarta.

Initially trained as a physician, Jacob went on to get degrees in anthropology in the United States and Holland. By the 1970s, his lab was the prime Indonesian location for anthropological studies. Soejono and Jacob both lived through the Japanese occupation of the Second World War. Soejono is famous for surviving an attempt to tear down a Japanese flag in front of gun-toting soldiers. Later, when a student army helped to drive out the Dutch colonial regime in 1949, Jacob was the students' voice on the lone radio station.

Both he and Soejono are national icons, with access to the presidential palace.

But Jacob is also an extremely proud individual, sensitive to hints of colonialism. Over the years, Indonesia has drawn international researchers to examine its geological and anthropological riches. Some of these visiting researchers have caused difficulties, at times taking too much credit for their discoveries. Jacob finds such behaviour insulting and has reacted negatively. Such dynamics have prompted some leading anthropologists to steer completely clear of Indonesia.

Somewhat scientifically isolated, the 75-year-old Jacob has continued to adhere to the theory that hominin species from multiple regions around the globe evolved into a single line that produced modern *H. sapiens*. In this theory, Neanderthals were direct

ancestors of modern humans, but most studies have shown that they were not related to *H. sapiens*. The 'multi-regional' theory has virtually no support among the world's leading anthropologists.

Instead, the prevailing view is that modern humans first evolved less than 200,000 years ago in eastern Africa, and spread across Europe, Asia and the available land bridges, becoming the dominant species. Reports of a human relative that existed 7 million years ago in Chad<sup>4</sup>, and a 160,000-year-old subspecies of *H. sapiens* in Ethiopia<sup>5</sup> have strengthened this 'out of Africa' hypothesis. The finding of *H. floresiensis* fails to support the multi-regional theory because there is no evidence that it contributed to the *H. sapiens* line. But Jacob and his allies remain unconvinced.

Their beliefs may even have affected Morwood's fieldwork. Following the discovery of LB1, the Indonesian Centre archaeologists excitedly sought more bones to ensure that the specimen wasn't just an individual with unusual characteristics. But last August, Soejono tried to pull the young Indonesian researchers out of Liang Bua cave. He claimed he wanted the team to help prepare for a scientific conference, but team members were not convinced by this explanation, and found the interference unhelpful.

In the end, Morwood's arguments to keep the researchers digging were successful, and the bones of a total of eight separate *H. floresiensis* individuals were unearthed. Buried at the same site was an unusual set of flaked stone tools<sup>6</sup>.

### Hand overs

In November, just after the publication of *H. floresiensis*' description, Soejono independently signed agreements to hand the specimens over to Jacob, who wanted to examine them himself. "We thought we would never see them again," says Morwood. Peter Brown, the lead author on the paper announcing the discovery, "worked furiously to take measurements and pictures before they were taken away", adds Morwood.

Jacob held on to the specimens for two months longer than was originally agreed, saying he needed more time. After repeated requests, all but two leg bones were returned on 23 February 2005 (see *Nature* 434, 5; 2005). Soejono says he was confused by the situation, caught between the desires of

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

Unearthing the hobbit's ancestor  
may help explain its small stature.

**"This callous treatment  
of hominin material of  
world importance is  
just sickening."**

— Michael Morwood

his scientific peer and those of the younger scientists. "My students were angry," Soejono laments.

Jacob also shared access to published and unpublished material with other researchers, including Alan Thorne, a long-time professional antagonist of Brown and a semi-retired palaeoanthropologist from the Australian National University in Canberra.

After their analysis, Jacob and Thorne publicly declared that *H. floresiensis* was not a new species. They believe the specimens are of a tribe of pygmies, one of which has the congenital disorder microcephaly — a disease that causes a smaller skull and brain. They hold to this position even after such a disorder was largely ruled out by imagery of the skull<sup>7</sup> published last month by US researchers who collaborated with the discovery team.

Putting such criticism behind them, the team this month turned its attention to the condition of some of the most important bones. The researchers believe that efforts to make casts at Jacob's lab, along with the trauma of transport, led to significant damage. Parts of the skull were pulled off, they say, and a jaw was

broken between the front two teeth, splintering an area crucial for future analysis. "There was no reason to make casts of such fragile bones," says Brown, noting that computerized X-ray images are just as useful. There was even an attempt to force pieces back together with glue, he says, and the pelvis of LB1 was crushed. "This callous treatment of hominin material of world importance is just sickening," says Morwood. Jacob declined a telephone interview with *Nature*, although he has denied damaging the fossils to others.

### Remains of the day

While the damage is assessed and further work is done on the bones, Morwood and his team are trying hard to look forwards to future discoveries.

So far, they are no closer to answering their original questions about who crossed the Wallace Line and when. Their excavations have shown that *H. floresiensis* was on Flores as early as 90,000 years ago<sup>6</sup>, and researchers have shown that other hominins were around at the same time. While *H. floresiensis* thrived, *H. sapiens* was living on nearby Borneo and *H. erectus* could be found in Java. But a clearer picture of the relationship between these different species is still elusive.



Cave spotting: Douglas Hobbs (left) and Michael Morwood crossing the Wallace Line on a ferry to Lombok to investigate possible dig sites (above).



Analysing DNA from the bones can help to determine a species' lineage. But first one has to find non-degraded genetic material, which the discovery team has not yet been able to do. A

team at the Max Plank Institute for Evolutionary Anthropology in Leipzig, Germany, is also looking for DNA, in a piece of rib supplied to them by Jacob. The experiments are currently under way.

In the meantime, Morwood and his team will spend the coming season digging in locations dictated by the monsoons. In February they began at Song Gupuh, which will stay dry no matter how hard it rains; they will move on as the monsoon recedes.

If Morwood's team finds more bones of *H. floresiensis* at Liang Bua, they will strengthen the published case. Multiple skulls with the distinctive small size will quieten even their most outspoken critics. But how many would they need to find? When asked face-to-face, Jacob hesitates for a moment, before saying: "Two or three? No. But if there were more, I might reconsider." He's on safe ground — finding such a trove of ancient skulls is basically unheard of.

Even finding good caves in which to look is proving tricky. After taking the ferry to Lombok, Morwood and fellow Australian Douglas Hobbs set off to follow a tip from an archaeologist friend. They found the cave he had spotted while on a surf trip. Saung Batu,

as it is called, is ideal for human habitation: spacious, dry and accessible. Inside, there is a rock used by some ancient folk for grinding seeds, and nearby rests a spear point. But the cave isn't suitable for a major dig. "The sediment doesn't look deep enough," proclaims Morwood.

In the months ahead, the team knows there may be more serious obstacles to its work than bad luck. The researchers must tread softly when challenging the greybeards of Indonesian science; no one wants to stir political animosity that might block exploration permits.

So far, the Australian team members have the backing of Tony Djubiantono, director of the Indonesian Centre for Archaeology, which issues permits. And the young Indonesian researchers are enthusiastic about their partnership with the Australians, grateful for the funds, training, patience and goodwill

brought to the project.

But these young Indonesians, eager to lead the team to the forefront internationally, must contend with their scientific elders. "It is not the Indonesian way" for the young to lead, Soejono has told them. Those researchers who know the potential for Indonesian anthropology hope that the way is changing. ■

Rex Dalton is *Nature's* US West Coast correspondent.

1. Brown, P. *et al. Nature* **431**, 1055–1061 (2004).
2. Morwood, M. J. *et al. Nature* **392**, 173–176 (1998).
3. Swisher, C. C. *et al. Science* **274**, 1870–1874 (1996).
4. Brunet, M. *Nature* **418**, 145–151 (2002).
5. White, T. D. *et al. Nature* **423**, 742–747 (2003).
6. Morwood, M. J. *et al. Nature* **431**, 1087–1091 (2004).
7. Falk, D. *et al. Science* published online, doi:10.1126/science.1109727 (2005).



# Immigration could ease climate-change impact

A modest proposal to allow the big gas-emitters to take their share of responsibility.

*Sir*—The recent UN World Conference on Disaster Reduction in Kobe, Japan, was dominated by discussion of the Indian Ocean tsunami (“Solo efforts hamper tsunami warning system” *Nature* **433**, 343; 2005). However, we must not forget the extreme vulnerability of small islands and low-lying coastal areas to sea-level rise caused by climate change.

The accumulation of greenhouse gases in the atmosphere during the past few hundred years is likely to result in a sea-level rise of up to half a metre, possibly more, by 2050 (R. J. Nicholls & J. A. Lowe, *Glob. Environ. Change*, in the press). Many of the affected countries do not have the resources to adopt protective measures such as sea walls and embankments, nor can they afford insurance. By the end of the century, millions are likely to have been driven from their homes by sea-level rise.

One of the ironies of climate change is that, although wealthy countries are

responsible for most of the accumulated greenhouse gases in the atmosphere, they will probably face less damage than poor countries. As climate negotiations ‘beyond Kyoto’ take shape, it is time to consider a framework wherein people living in areas likely to be rendered uninhabitable by climate change would have the early option of migrating elsewhere — specifically to those countries that are largely responsible.

The number of vulnerable ‘climate-change exiles’ received by a host country would be in approximate proportion to that country’s cumulative greenhouse-gas emissions. Estimates suggest that roughly 50 million to 200 million people will be displaced by the 2080s, owing to the direct impacts of climate change under a plausible range of emissions scenarios (R. J. Nicholls *Glob. Environ. Change* **14**, 69–86; 2004). Assuming that all these climate-change exiles are absorbed by the top ten ‘emitter’ countries, new annual immigrants would

range from a few thousand for the Czech Republic to about three-quarters of a million for the United States.

Once the basic principle (which is consistent with Articles 1 and 4.8 of the United Nations Framework Convention on Climate Change) is accepted, there will be several ways to determine who should be considered for immigration benefits, which countries should bear the costs of immigration, and what institutional and political mechanisms are needed in order to minimize the risks of a massive refugee crisis as climate impacts become more severe.

Figuring out an international strategy to address the needs of these climate-change exiles will be easier now than later.

**Sujatha Byravan\*, Sudhir Chella Rajan†**

\*Council for Responsible Genetics, 5 Upland Road, Cambridge, Massachusetts 02140, USA

†Tellus Institute, 11 Arlington Street, Boston, Massachusetts 02116, USA

## Consumer law is used to attack climate findings

*Sir*—Your News story “Salt sellers challenge US health agency using data-quality act” (*Nature* **433**, 671; 2005), suggests that a lawsuit filed by the Salt Institute represents “the first time that a petitioner has actually sued under the Data Quality Act”. This is not the case, although the ruling that the Salt Institute is appealing was the first verdict given in a case brought under the federal Data Quality Act.

In August 2003 a conservative advocacy group, the Competitive Enterprise Institute, filed a suit claiming violations of the act against President George Bush and John Marburger, director of the White House Office of Science and Technology Policy (OSTP). The most important claim was that reports by the US National Assessment of the potential consequences of climate change (USNA) used results from two different global climate models to construct scenarios of future climate (see National Assessment Synthesis Team, *Climate Change Impacts on the United States*, Cambridge Univ. Press, Cambridge, 2000) — which therefore violated the act because one of the two scenarios had to be in error.

In November 2003, just as the Department of Justice was preparing to file its response, the parties accepted a ruling of “dismissal with prejudice”, meaning that

the lawsuit could not be refiled. Although the Department of Justice did not release its brief, it had apparently made a strong argument against the absurd notion that projections of the future must be proven accurate in advance.

The OSTP nonetheless ordered that a notice be added to USNA web pages, indicating that the reports “were not subjected to OSTP’s Information Quality Act Guidelines”. This implies that the USNA report was not properly reviewed and would not meet the OSTP guidelines. This is misleading at best, as the report was subjected to a four-stage review that was more comprehensive than called for by the act. In addition, the OSTP guidelines did not exist or apply at the time that the USNA was released.

Attempts by the USNA’s lead authors and contributors (including myself) to get this notice removed or modified have failed, leaving in place an unfair criticism of the assessment’s widely accepted findings.

**Michael C. MacCracken**

Climate Institute, 1785 Massachusetts Avenue NW, Washington DC 20036, USA

## That chemist pose is a classic because we do it

*Sir*—I note with some surprise that Philip Ball, in his *Science in Culture* article “What’s in the flask?” (*Nature* **433**, 17; 2005), says that gazing into a flask held

aloft — as in the stereotypical image — “is not what real chemists spend their time doing”. As a synthetic organic chemist, I would like to point out that this is exactly what real chemists do.

My research group regularly observes the initiation of crystallization or precipitation in this fashion. Magnification through the curved walls of the flask, with the brightly lit background of the ceiling lighting, lets one see the changes in the morphology of small particles as they form (often at the liquid surface), while trying to avoid solutes crashing out of solution unselectively.

As chemistry is so often the glue connecting biology, physics and medicine in this interdisciplinary age, we can only hope that more people will discover what chemists actually do.

**Piers R. J. Gaffney**

Department of Chemistry, Imperial College, South Kensington Campus, Exhibition Road, London SW7 2AZ, UK

**Erratum** In David M. Wilkinson’s Correspondence letter “Brown knew particles were smaller than pollen” (*Nature* **434**, 137; 2005), the sentence “The microscopic particles involved in the characteristic jiggling dance Brown described were much smaller clay particles” should not have contained the word “clay”, which was added in error during the editing process. The particles Brown saw were inside the pollen grains.

## correspondence

Contributions to Correspondence may be submitted to [corres@nature.com](mailto:corres@nature.com). They should be no longer than 500 words, and ideally shorter.



# The never-ending story

A guide to the biggest idea in the Universe: infinity.

## The Infinite Book: A Short Guide to the Boundless, Timeless and Endless

by John D. Barrow

Jonathan Cape: 2005. 328 pp. £17.99. To be published in the US by Pantheon Books (\$26).

Simon Singh

John Barrow is a wide-ranging author. A few years ago he wrote *The Book of Nothing* (Jonathan Cape, 2000), an exploration of zero, and now he has published *The Infinite Book*. Again he tackles a subject that has been previously explored by other popularizers of science, but again he brings his charm and wit to bear to provide an account that is highly engaging and enriched with numerous literary references and dozens of illustrations and photographs.

The opening chapters introduce readers to the concept of infinity through the eyes of the various philosophers who tackled the subject through the centuries. Their conflicting views about the nature and existence of infinity are slightly baffling, but the situation is rescued by the nineteenth-century mathematicians who were able to make sense of the infinite realm. What emerges is a series of staggeringly brilliant and beautiful ideas, whose logic and power seem to defy common sense.

For example, are all infinities equal? There are an infinite number of numbers and an infinite number of even numbers, but is the former infinity double the size of the latter, or does infinity have hidden subtleties? The Infinite Hotel is the classic device for addressing this question. Attributed to the German mathematician David Hilbert, the Infinite Hotel is a highly successful enterprise and its infinite number of rooms are fully booked. The situation seems bleak when a new guest turns up at reception, but the resourceful hotelier, comes up with a solution. He asks everybody to move up to the next room — the occupant of room 1 moves to room 2, the occupant of room 2 moves to room 3, and so on. This means that everybody still has a room, but room 1 is now empty and is available for the new guest.

This merely shows that infinity plus one equals infinity. But what if an infinite number of new guests arrive at reception? The hotelier asks the current guests to move to the room with the number that is double their current room — the occupant of room 1 moves to room 2, the occupant of room 2 moves to room 4, and so on. This means that everybody still has a room, but all the odd-numbered rooms are now empty and available for the new guests. So infinity plus

infinity equals infinity, and hence the number of even numbers is equal to the number of all numbers.

Barrow explains other arguments that show that the number of whole numbers is equal to the number of fractions, but that the number of decimals is larger and represents a different scale of infinity. He also tells the tragic story of Georg Cantor, a German mathematician who developed many of the fundamental ideas relating to infinity. Cantor was heavily criticized by the 'finitists', such as the influential Leopold Kronecker, who successfully vetoed Cantor's applications for senior academic posts and blocked the publication of his papers.

This constant persecution led to a series of nervous breakdowns and spells of severe depression for Cantor, who gradually moved away from mathematics and spent increasing amounts of time with philosophers and theologians. Eventually Cantor did receive some acclaim from mathematicians around Europe, but in his German homeland he was largely ignored. In 1908 he complained that his German colleagues "do not seem to know me, even though I have lived and worked among them for 52 years".

Some chapters move beyond mathematics and deal with the infinitely large in the

context of the Universe and the infinitely small with respect to particle physics. Barrow also ventures even further afield, touching on theological issues and examining eternity and the notion of elixirs for immortality.

Occasionally the material feels familiar, but Barrow is generally able to introduce novel twists and turns, and presents standard material in refreshing ways. For example, it is well known that the density at the centre of a black hole is supposed to be infinite, but few people realize that the average density of a large black hole, such as the one at the centre of a galaxy, could be less than the density of air. This is because the Schwarzschild radius of a giant black hole, which defines the point of no return, is so vast.

Familiar characters, such as Hilbert, are also fleshed out in new ways. We learn that one of Hilbert's students committed suicide when he failed to solve a particular mathematical problem. Hilbert was asked to speak at the funeral, so he stood at the graveside and matter-of-factly explained that the problem was not particularly difficult and that the young man had merely failed to look at it in the right way.

Even an old chestnut like the Infinite Hotel feels fresh. This results in an index that lists Fawley Towers next to Fermat's last

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

Seeing the big picture: the patterns in Islamic tile designs can be repeated indefinitely.

theorem — surely the only time that these two items have been neighbours in the same book — and this serves to indicate the book's quirkiness. ■

*Simon Singh is the author of Big Bang (Fourth Estate, 2004), a history of cosmology.*

## Wine with a deep flavour

### **The Winemaker's Dance: Exploring Terroir in the Napa Valley**

by Jonathan Swinchatt & David G. Howell  
University of California Press, 2004. 243 pp.  
\$35.95, £22.95

**George W. Moore**

The taste of a wine depends on more than just the variety of the grape from which it is made. Winemakers and wine connoisseurs are concerned with *terroir* (pronounced 'tair-wahr'), which refers to everything else that controls the flavour of wine: climate, subsoil and even the manipulations done in the winery.

Many winemakers believe that the taste of wine comes from the soil that supports the vines. But the soil is only about a metre thick, whereas the roots of the vines extend down about 10 metres or more. The soil is in equilibrium with the local climate: its clay and organic components smooth out the delivery of water to the plants, and cation-exchange processes supply major nutrients. But the minor compounds that control the taste come from the much more voluminous subsoil. The bedrock slowly weathers under the action of water, carbon dioxide and organic acids. This process delivers a continuous flux of subsoil-dependent compounds to the vines, and is a key component of the local *terroir*. In *The Winemaker's Dance*, geologists Jonathan Swinchatt and David Howell evaluate every aspect of the *terroir* of a small but well known winemaking region and place their findings in a global context.

The Napa Valley, northeast of San Francisco, is in many ways similar to the famous Bordeaux region of France. They are at similar latitudes on the west coasts of continents, and principally grow the warm-climate grape variety Cabernet Sauvignon. They are both about 50 kilometres from their respective oceans, and large estuaries — San Francisco Bay at Napa and the Gironde at Bordeaux — provide the cool night-time temperatures that preserve the acidity of their premium wines.

The most expensive Napa Valley wines grow on alluvial fans (known locally as 'benches'), which are piles of gravel and interbedded silt that have streamed down off the flanking mountains. At Bordeaux, the counterparts to Napa's alluvial fans are



**I'm getting something earthy: differences in bedrock can affect both soil colour and a wine's flavour.**

the *graves* (gravel mounds, again interbedded with silt) formed by streams during ice-age low stands of the estuary.

In an effort to link the taste of wine to the subsoil geology, the authors studied Diamond Creek Vineyards, where the vines are planted on upland slopes at the north end of Napa Valley. The 8-hectare property is cut through by a network of faults that juxtaposes three dissimilar geologic blocks. During vineyard development, the owner recognized three strongly contrasting soil colours, and used them to divide the property into three vineyards. These were planted in 1968 with Cabernet Sauvignon cuttings from two first-growth vineyards at Bordeaux.

Throughout its history, Diamond Creek has had three vineyard-designated bottlings. Similar winemaking practices are used for each but the wines consistently taste different. Wine from Red Rock Terrace (which has a basaltic lava subsoil) has velvety tannins with rich cherry and blackcurrant flavours; that from Volcanic Hill (volcanic-ash-bearing sandstone) has tannins that are youthfully firm but long-lived with berry and smoky flavours; and wine from Gravelly Meadow (pebbly stream sediment) has firm tannins with herb, blackberry and earthy flavours.

Two major factors affect the subtleties of wine taste. The first is the 'character' of the wine — the grape variety plus all the elements of its *terroir* — and the second is the physiology of the people doing the tasting. Tastes vary: one person's *crème de la crème* is another person's plonk

To investigate the effect of local *terroir* on the palettes of ordinary consumers, Swinchatt and Howell conducted a blind tasting in 2002, serving eight premium Napa Valley wines with a meal. Most participants ranked the wines differently when they first tasted them to when they tried them again with a meal. When asked which of the eight they preferred overall, the votes were spread quite evenly. In an extreme case of differential preference, one person found the aroma of

one wine to be particularly offensive, whereas another said of it: "I'd like to have a bowl of that beside my bed, just to smell."

This variation in personal preference is in marked contrast to the 'expert' opinion of wine-ranking services, which tend to favour blockbuster wines that are extremely intense, particularly on first tasting. This has led to winemakers letting their grapes hang as long as possible on the vines, and extending the time that the wines macerate on the grape skins. Swinchatt and Howell deplore the homogenization that this is causing to the taste of the world's wines. They recommend that consumers ignore the wine-ranking services, seek out diversity, and savour it. ■

*George W. Moore is in the Department of Geosciences, Oregon State University, Corvallis, Oregon 97331, USA.*

## The quantum Universe

### **Science and Ultimate Reality: Quantum Theory, Cosmology and Complexity**

edited by John D. Barrow, Paul C. W. Davies & Charles L. Harper, Jr  
Cambridge University Press, 2004. 742 pp.  
£40, \$60

**Frank Close**

In the centenary year of Einstein's seminal contributions to human culture, *Science and Ultimate Reality* celebrates the 90th birthday of John Wheeler, who took two of Einstein's ideas and created new branches of science. The physics of the twentieth century was built on the twin pillars of Einstein's general theory of relativity and quantum mechanics. The former describes the macroscopic Universe of space-time and gravitation, and the latter is a theory of matter and radiation at subatomic dimensions. These two different



## Museum

## Waxing and waning

The art of using wax models to demonstrate anatomy in three dimensions reached its zenith during the eighteenth century in the work of Ercole Lelli, in Bologna, and Clemente Susini, in Florence.

The most renowned collections of wax anatomical models are displayed in museums in these two Italian cities. But the Orfila Museum at the University of Paris V in France also holds important examples, including later works such as this 1847 model by one of the university's own anatomists, which was created for the museum's opening.

If you want to see the Paris collection, you may need to be quick, however. Unlike the Italian museums, the Orfila has struggled to survive, despite having been listed as a French National Heritage treasure in 1991. The university is reclaiming the exhibition rooms, and the museum's entire collection will be packed away into a basement later this year.

The Paris museum's collection has

always been subject to dramatic swings in fortune. In the Second World War, for example, many of the wax models made by anatomist Jean-Baptiste Laumonier at the beginning of the nineteenth century were used to make candles, and only a few hundred remain in the collection today.

Despite this, the museum still hosts almost 6,000 historical items. These include, in addition to the wax models, anatomical models using other early twentieth-century techniques and materials. There are also casts of brains commissioned by neuropathologist Paul Broca — best remembered for his discovery of the part of the brain that controls speech — and an extensive collection of anthropological specimens, including Neanderthal and *Australopithecus* skulls and bones.

The museum can be visited by appointment.

**Achim Schneider**

♦ [www.biomedicale.univ-paris5.fr/anat](http://www.biomedicale.univ-paris5.fr/anat)



theories seem to be incompatible, but Wheeler insisted that, at some level, quantum effects must have a meaningful impact on gravitational physics. And because Einstein's theory intimately links gravitational physics with geometry, the net result must be some sort of quantum space-time dynamics. It was using this reasoning that Wheeler predicted in 1957 the existence of space-time foam, one of many of his ideas that survive to this day.

Wheeler is popularly known for having

coined the phrase 'black hole', and professionally for his wide range of profound contributions in physics, ranging from the very large to the smallest scales of size. Gravitation was reborn, with Wheeler's influence, as a mainstream branch of science, leading to the explosive growth in astrophysics and cosmology that we see today. At the other extreme of scale, he developed the theory of nuclear rotation, with Edward Teller, and the fundamentals of electrodynamics, with Richard Feynman. And it was Wheeler's seminal question that led to the perception of the positron as an electron travelling backwards in time.

His 90th birthday was celebrated with a symposium attended by a metaphorical galaxy of scientific stars, and this book records the event. Do not be put off by its length: at over 700 pages this is still only about ten pages per year of active research, and this is one of those rare volumes where quantity is matched by quality.

With such a variety of contributions — ranging from quantum reality (by Freeman Dyson and others), big questions in cosmology (including articles by Andreas Albrecht, John D. Barrow and Andrei Linde), higher dimensions (Lisa Randall) and the emergence of life (George Ellis) — how can a review do justice to them all? Perhaps I should focus on the man himself, aptly summarized in a readable and insightful opening chapter by Paul Davies.

Wheeler's gift has been in asking questions that are, in the modern parlance, 'outside the box'. While most of us rack our brains trying to determine and understand the implications of nature's laws, Wheeler

wondered whether the very concept of physical laws might be an emergent property. Could law-like behaviour emerge stepwise from the ferment of the Big Bang, instead of being mysteriously and immutably imprinted on the Universe at the instant of its birth? As Davies says: "Wheeler was breaking a 400 year old scientific tradition of regarding Nature as subject to eternal laws."

As students we learn quantum mechanics, and as professionals we apply it with varying levels of unease as to what it actually means, but Wheeler had a singular attitude. He insistently asked: "How come the quantum?" Why is the world quantum mechanical? What would happen if we made small changes to quantum mechanics? The preface asks the same sort of questions as Wheeler: "Could it be that quantum mechanics is the simplest mechanics consistent with the existence of conscious beings?" Or does it "optimise the information processing power of the universe"? These are profound questions, as yet unanswered but extensively and profoundly discussed, and some sense of the debate can be found in this book.

Wheeler was the inspiration behind the most extreme connection of all: that quantum mechanics, a theory of subatomic dimensions, can be applied to cosmology, the largest system of all. In response to Einstein's question "Did God have any choice in the nature of his creation?", Wheeler has suggested that there are no truly fixed fundamental laws of physics at all. He was a remarkable man and this is a remarkable volume. ■

Frank Close is at the Rudolf Peierls Centre for Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK.

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

A sign of success: John Wheeler is credited with coining the phrase 'black hole'.

# Schrödinger's mousetrap

## Part 10: The trap is sprung.

Henry Gee

Lister had finally lost the battle. Coping with disputatious, driven scientists was far worse than dealing with gun fights on a Saturday night. And trying to imagine what entangled photons looked like had pulled his mind in more directions than he could cope with. In any case, the nicotine patches had failed. He pulled a crumpled pack of Marlboro Reds from his pocket, yanked one free and lit up. The tang of fresh tobacco cast Nigel Lorimer back to his first day at *Nature* many years before, when he was employed by an editor whose reputation for shrewdness equalled his notoriety as a chain-smoker.

"Want one?" Lister asked the suspects gathered in the seminar room. Tony Trotman put up his hand without a second thought. Wilfred de Bruijn hesitated before accepting. Anything, Lister thought, to steady their nerves. His, too.

The crash course in physics, although challenging, had set him on the right path. "I guess you know why I've brought you all here," he began. Petra Pruszczyński looked defiant; Veronique Dubois, impassive; de Bruijn, quietly seething. Fenton Baumgarden? Jirong Feng? Zen surf dudes, both. Ludmilla Shlomiuka had a thousand-yard stare. Trotman withdrew to the back and Lorimer was all attention.

"The reason," continued Lister, "is that each one of you had reason to loathe Rufus Jaeger — perhaps reason enough to kill him. Rufus had you by the balls, Mr Trotman, after that fire you started. And he snatched a Nobel from you, Professor Baumgarden." The Californian said nothing. "Neither of you has a decent alibi for the coffee-break immediately before Rufus's circus act, either. Professor Baumgarden was — get this — meditating, and you, Mr Trotman, say you went for a smoke.

"Those of you who were mingling among the coffee and biscuits aren't off the hook just yet, though. You all still had motive, and there's more than one way to skin a quantum cat." Nobody laughed.

"You're a dedicated editor, Nigel." Lorimer looked up, startled. "But Rufus burst your bubble. And you, Professor Pruszczyński," he almost spat, "hated Rufus Jaeger and all he stood for.

"He stole some of your best ideas, Professor Dubois. Would you kill him, just to get even?" Dubois' lips compressed to a sere line, the question was left hanging in space. Lister pressed home. "You spent the coffee break with Professor Pruszczyński, 'collaborating'



## ACE DETECTIVE!!



as you put it. Nobody else can cover your alibis. Collaborating on murder, maybe?

"And Dr Shlomiuka loves everyone, but she has history. Or rather, her mother did. Back in the old country. You found out soon enough that Rufus was not your father but you have been obsessed with your mother's past. Am I right?" Only then did Shlomiuka look up, jolted from her reverie.

"Mr Feng, sent packing in disgrace — again by Rufus — and you, Wilf, forever turned down for promotion by Rufus, who steals not just your ideas but your best girl. And where did you go after being angered to boiling point during the coffee break?" De Bruijn looked broken and, like his spent cigarette, as if all the fire had gone out of him.

"It turns out, apparently, that a prism in Rufus's contraption was switched," Lister went on. "In its place we found one made of a substance with what's called a negative refractive index. The result was that a power-

ful laser beam drilled through Rufus's head. I believe it's called 'negadex.'" Pruszczyński gasped.

"The prism had to be switched during the coffee break before the presentation — only Ludmilla, Jirong and Nigel have good alibis. The rest of you, as I've said ... don't." Pruszczyński looked like she was about to explode.

"But let's step back a little," continued the detective (at last, he was beginning to enjoy himself). "The negadex is a red herring, for when the prism was switched, Rufus Jaeger was already dead." His pause for effect was, well, effective. He lit another Marlboro.

His next words broke what seemed to be a hostile and active silence. "He was killed by a lethal drugs overdose. Now, I don't know much about physics, but narcotics are something I do know about."

Confusion. Pruszczyński, de Bruijn and Lorimer began to talk at once. Lister pressed

CHRISTIAN DAKIN



on regardless — all fell silent, to hear where the axe would fall.

"If Rufus was already sitting dead in his chair, the prism could have been switched during the demonstration. Could you have poisoned him, Professor Baumgarden? You broke a confidence to tell me about negadex," Lorimer looked angrily at Baumgarden, "diverting me from the real cause of death. And you were chairing the session. Perhaps you dosed his glass of water? And what was your real reason for inviting Rufus to give the plenary lecture?"

"Now just wait a goddamned minute..." began Baumgarden, visibly agitated for the second time, but Lister waved him down. Forensics had said Jaeger's glass on the podium was clean. "Petra and Veronique hated Rufus enough, and have no alibi, either — but they had no opportunity to slip Rufus any poison." Except with their eyes, he thought, meeting Pruszczyński's enraged stare.

"Nigel and Jirong are in the clear," Lister continued, "both having alibis for the coffee break and sitting a safe distance away from the podium. And are Tony, Wilf and Ludmilla likely to have known about negadex? Tony and Wilf, no. But, you see, it all comes back to Dr Shlomiuka."

The others shrank back from her, involuntarily, as if she bore a curse. "She knew Rufus very well. Just like she knows all of you. Dear Ludmilla, how she tries to please everyone. Dear, poor Ludmilla. Accent on the 'poor'. The expense of keeping her husband Dmitri alive! The pressure! The stress! How well she has learned her pharmacology! She kept up her acquaintance with Jirong, and softly but determinedly put him under pressure, making him feel she needed his help and that he had some favours to return.

"I am guessing that now and then he would provide her with various bits of confidential information. Eventually, he even handed her a piece of negadex from Gdansk, all the while innocently believing that Ludmilla was just being curious. Another miscalculation, my dear Jirong."

Pruszczyński turned to Feng with a shaft of pure, icy venom. Feng looked away as if slapped. His first error had been a well-publicized mistake; his second, well, would Petra count that in her 10%? But still Lister rolled along.

"Being Ludmilla, she at first only wanted the negadex to praise her mentor, not to bury him." Damn. Must stop these literary allusions. He'd get laughed off the force.

"But then Ludmilla got a q-mail offering her money to kill Rufus. A lot of money. Enough to keep Dmitri comfortably numb." Another mental note. Stay off the vintage rock allusions, too, especially those floydian slips. "She quickly real-

ized that her piece of negadex could be invaluable in such an enterprise. So she poisoned Rufus with her husband's drugs during the coffee break, when she had a perfect alibi — just a little, the exact right amount in a chocolate biscuit she knew Jaeger couldn't resist — so that some way into the presentation, he would expire. The timing was crucial, but she knew precisely how these drugs would work, and when Rufus was stiff in his chair, Ludmilla could switch the prisms so that nobody would notice, confident that only those who knew about negadex could be blamed."

This time it was Trotman's face contorted in fury — how could anyone tamper with his machinery?

"But why would I want to kill Rufus?" protested Shlomiuka. "I liked him!" She burst into sobs — too well-practised, thought Lister, but he stepped down from the dais and sat next to her, reassuring.

"It's as Oscar Wilde wrote it," he almost whispered — listening to himself in abject horror. "We always kill the thing we love. And then there was the money. The money you needed for Dmitri."

"But I never..."

"It was that q-mail from de Bruijn that finally drove you to murder."

Shlomiuka looked puzzled.

"What?" shrieked de Bruijn, "all I did was complain about him stealing my work!"

"Ah yes," said Lister, meeting de Bruijn's incredulous eyes. "What you tried to send was this..." He waved to the back of the room. One of two impassive policemen, guarding the door, dimmed the lights. The other switched on a projector. Time for his own surprise presentation.

A message flashed up on the whiteboard, the typeface ragged under magnification.

*Ludmilla, hi. I think you should be careful about your relationship with RJ. He is a liar and a cheat. I will give you an example. He has just accepted a prize of 100,000 Euros for the paper on parallel quantum codes which is entirely my work. I realize now that I should have asked the journal to eliminate Jaeger from the list of authors. Wilf.*

Lister spoke out of the shadows, his form only partly illuminated by the incriminating quanta. "But thanks to the laws of physics, Wilf, that wasn't the message that got

through. As a way of testing her own, rival system, Professor Dubois intercepted the message, destroying most of the text." Dubois looked horrified as the awful truth superposed itself on her mind. "This," said Lister, "was what Ludmilla actually received. And knowing Wilf and his hatred for their boss quite well, she never doubted that the message was genuine."

The old slide dissolved; a new one revolved into view.

*Ludmilla I will give you 100,000 Euros to eliminate Jaeger. Wilf.*

Shlomiuka stood up: "You mean there never was any money on offer? That I killed Rufus because of a silly mis..." The mouse-trap sprang shut.

One of the silent policemen switched on the lights. The other led the sobbing and unresisting Shlomiuka outside to the waiting police car. Job done.

"Thank you, ladies and gentlemen," said Lister. "You are free to go — for the present." He left the room, leaving the door ajar.

The demonstration had been a triumph, almost matching Jaeger's own inflated view of his own showmanship. Pruszczyński was quite right, of course: showmanship was what it was. But there was a serious point to it all — something dangerous, too dangerous to reveal to the conference organizers.

Doing experiments in the quantum macroscopic realm could have unforeseen effects on the experimenters, too, and the other observers, including the audience. Just look at it from the viewpoint of Schrödinger's cat — the audience, too, existed in a superposition of states. Jaeger believed that his line of work might even establish a connection between quantum mechanics and the einsteinian world of space-time, such that reality itself could be altered — or, he thought, our perception of reality, which amounted to the same thing.

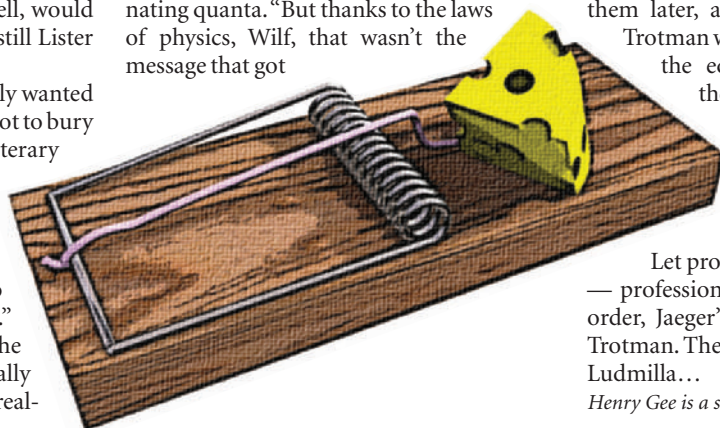
He kept these views to himself, of course — wisely, he thought. For now, the showmanship was the thing.

After the session, Baumgarden, as chair, took him out to lunch: just the two of them, plus Lorimer, of course (he just couldn't shake the guy free). Ludmilla would join them later, after calling Dmitri at home.

Trotman would be occupied dismantling the equipment. At the restaurant they found themselves next to a table at which Pruszczyński, Dubois, Feng and de Bruijn were already seated. At Lorimer's suggestion they pulled the tables together.

Let professional rivalries stay just that — professional. Just as they were about to order, Jaeger's mobile phone went. It was Trotman. There had been a terrible accident. Ludmilla...

Henry Gee is a senior editor at Nature.



# Hotheaded healer

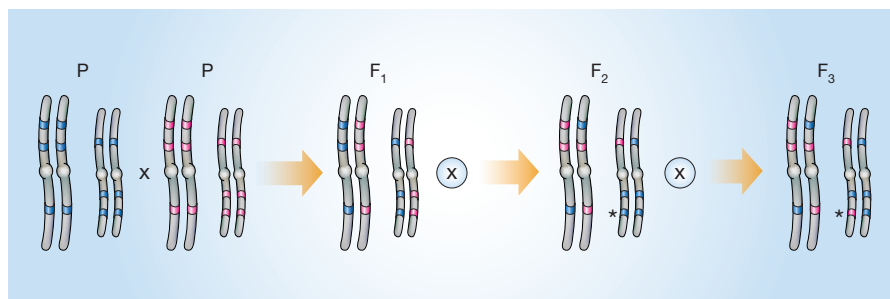
Detlef Weigel and Gerd Jürgens

A previously unknown way of reversing genome-wide sequence changes in DNA has been revealed by an analysis of plants carrying mutations in a gene called *HOTHEAD*. The mechanism remains a mystery.

Variety is the spice of life. This adage applies as much to genetic material as to anything else — if DNA were static, there would be no change over time and no evolution. Everything in moderation, though, to use another proverb. Most random changes in DNA (mutations) are harmless, because much of our DNA has no essential function. Nevertheless, of all possible mutations, the ones that are detrimental far outnumber those that might have a positive effect. Unwanted changes in DNA, which occur quite often, can have disastrous consequences and lead to all sorts of diseases. How, then, does the genome heal itself? Lolle and colleagues<sup>1</sup> have just had the first glimpse of a novel mechanism that does just that. They report their findings on page 505 of this issue.

The genome has sophisticated tools for correcting mutations every time a cell replicates its DNA in preparation for cell division<sup>2</sup>. This machinery, which discovers incorrect copying or damage in the DNA, is essential for avoiding the accumulation of unwanted mutations during our lifetime. There is nothing mysterious about this process, as it uses the intact DNA strand as a template for correcting mistakes on the broken or damaged strand. But can the genome also detect mutations that have become 'fixed', where the DNA sequence on both strands has changed? Reversion of fixed mutations is less unusual than one might suppose, but it generally relies on the original DNA still being present. For example, transposons, which are essentially parasitic sequences of DNA, may insert into a gene and disrupt its activity. But if the transposon is precisely excised, and the host DNA stuck back together, the gene reverts to its normal state.

More intriguing are cases in which fixed changes revert to the original sequence without an apparent template. For example, in microbes, revertants that confer some sort of physiological response can be easily selected from large numbers of individuals. Thus, although mutations back to the original sequence are rare, on the order of one in a billion per site and per generation, the large population size makes the isolation of revertants possible. There is even evidence that stress increases the frequency at which revertants occur<sup>3</sup>. Amazingly, precise reversion events have also been described in humans<sup>4</sup>. Unfortunately, it is unclear whether these



**Figure 1** Reversion in *hothead* plants. Two parents (P) with different DNA sequence variants at several positions (colour-coded) in the genome are crossed. In subsequent generations (F<sub>1</sub> and F<sub>2</sub>), plants are allowed to self-fertilize (indicated by a circled cross). In the F<sub>3</sub> generation, there is reversion (asterisk) to a sequence originally present in the grandparent, but not in the immediate parent. This has been observed for several sites at surprisingly high frequency in the *hothead* mutant.

events are more frequent than expected by chance.

By contrast, there is no doubt that the reversion rates observed by Lolle and colleagues<sup>1</sup> in the plant *Arabidopsis* cannot be explained by random mutations. The authors' spectacular discovery started with a mutant called *hothead*, in which various organs are fused. What made *hothead* special was not its morphological defects but the finding that several independent mutant strains yielded apparently normal progeny at a high frequency (a few per cent). Lolle and colleagues show that this is due to precise reversion that restores the original DNA sequence. They ruled out several trivial explanations: the reversion is not due simply to a drastic increase in mutation rate, and it could not have been caused by gene conversion, where a related gene from elsewhere in the genome is used as a template. Nor is the reversion specific to the *HOTHEAD* gene or to harmful mutations, as the authors show by using hybrids carrying innocuous sequence changes at several other sites in the genome (Fig. 1). They find that *hothead* mutant progeny at later generations can recover DNA variants that have come from one of their great-grandparents, even if their immediate parent did not contain the variant.

Why does reversion occur so frequently in *hothead* mutants? The *HOTHEAD* gene itself does not seem to encode a protein with an obvious connection to DNA repair<sup>5</sup>, so the authors speculate that the high reversion frequency may indicate a response to some sort of elevated metabolic stress. Another fascinating question is what template the cell uses to restore the original DNA sequence.

From the authors' experiments it seems unlikely that the template is a cache of extra-chromosomal DNA. Rather, it may well be an unknown species of RNA. This is not unreasonable, as DNA and RNA can be copied back and forth precisely<sup>2</sup>, and RNA molecules can, for example, guide dramatic DNA sequence changes in some organisms<sup>6</sup>. An obvious candidate would be messenger RNA, which is an intermediate molecule involved when the DNA sequence is being translated into protein. However, the sequence reversions observed by Lolle and colleagues are not restricted to those parts of the genome that are normally copied into messenger RNA.

Many experiments to explore this mechanism of unorthodox inheritance, which can skip several generations, come to mind. But it seems a safe bet that the isolation of secondary mutants that have lost the ability to produce healthy progeny will provide helpful clues.

Detlef Weigel is in the Department of Molecular Biology, Max Planck Institute for Developmental Biology, D-72076 Tübingen, Germany.

e-mail: weigel@weigelworld.org

Gerd Jürgens is at the Centre for Molecular Biology of Plants (ZMBP), University of Tübingen, D-72076 Tübingen, Germany.

e-mail: gerd.juergens@zmbp.uni-tuebingen.de

1. Lolle, S. J., Victor, J. L., Young, J. M. & Pruitt, R. E. *Nature* **434**, 505–509 (2005).
2. Watson, J. D., Losick, R., Bell, S. P., Gann, A. & Levine, M. *Molecular Biology of the Gene* (Benjamin Cummings, San Francisco, 2003).
3. Rosenberg, S. M. *Nature Rev. Genet.* **2**, 504–515 (2001).
4. Hirschhorn, R. J. *Med. Genet.* **40**, 721–728 (2003).
5. Krolkowski, K. A., Victor, J. L., Wagler, T. N., Lolle, S. J. & Pruitt, R. E. *Plant J.* **35**, 501–511 (2003).
6. Mochizuki, K. & Gorovsky, M. A. *Genes Dev.* **18**, 2068–2073 (2004).



## Genomics

## Frontiers of gene function

Sean M. O'Rourke and Bruce Bowerman

The technique of RNA interference continues to pay dividends. The latest application of the method to the nematode worm adds detail to the list of genes known to function in the early stages of development.

Genome sequencing projects have begun to provide a comprehensive picture of the hereditary information required for various forms of life. But what do all the genes so identified do to warrant their existence? In the unicellular yeast *Saccharomyces cerevisiae*, for example, all predicted genes have been deleted, and the resulting strains subjected to a battery of analyses. These genome-wide studies have shown that only about 19% of yeast genes are involved in essential processes<sup>1</sup>.

In multicellular organisms, deciphering gene function on a genome-wide scale is a much tougher prospect. Here, one of the main laboratory subjects is the nematode worm, *Caenorhabditis elegans* (Fig. 1). On page 462 of this issue Sönnichsen *et al.*<sup>2</sup> report their testing of nearly every *C. elegans* gene for functional requirements during embryo formation and larval development. They report that 1,668 genes display a visible phenotype, or defect, when reduced in function. A surprisingly small number (661) are essential for distinct cell processes, and embryonic viability, during the first two embryonic cell divisions.

In 1998, the release of the *C. elegans* genome sequence provided the first complete catalogue of the predicted gene content of a multicellular organism<sup>3</sup>. These data have allowed bioinformatic comparisons with other genomes. For example, sequence analyses reveal that about 50% of human genes have homologues in yeast, fruitflies or worms<sup>4</sup>. Also, so-called microarrays — glass slides with thousands of spots, each containing many copies of DNA specific for one gene — have allowed the rapid and systematic profiling of gene expression in numerous environmental conditions, at different developmental stages, and following genetic perturbations<sup>5,6</sup>. Next came methods to understand gene function in relation to the phenotypes produced when genes are eliminated.

Most recently, the remarkable phenomenon of RNA interference (RNAi) — the ability of double-stranded RNA (dsRNA) to interfere with the production of the corresponding gene product<sup>7</sup> — has allowed worm researchers to 'knock down' the expression of any given gene. RNAi was first applied by microinjecting dsRNA into worms<sup>8</sup>. More recently, it was shown that feeding worms dsRNA produced in bacteria (*C. elegans*' favourite meal) also reduces gene

function<sup>9</sup>, leading to the first genome-wide tests for gene function in *C. elegans*<sup>10,11</sup>.

These feeding RNAi studies have revealed phenotypes for about 10% of the worm's genes, with half of these (929) resulting in death of the embryo when they are reduced in function. But the phenotypes described in this first genome-wide study were divided into very broad categories (for example 'embryonic lethal', 'sterile', 'larval arrest' and so on). Although this information, and the reusable feeding RNAi library it generated, are extremely useful, more specific phenotypic descriptors were not provided. For example, embryonic lethality can result from any number of defects in just the first embryonic cell division.

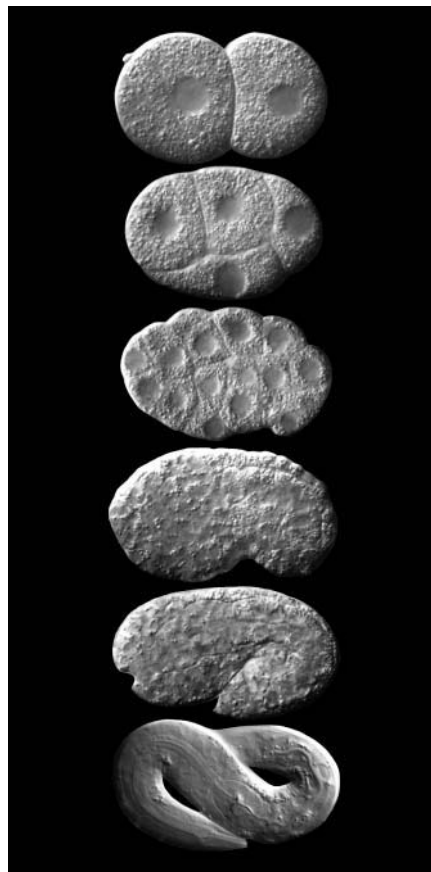


Figure 1 From embryo to organism. The development of *C. elegans* from the two-cell stage to the fully formed worm, just before hatching. Sönnichsen and colleagues' microscopic analysis during the first two cell divisions has allowed them to predict functions for previously uncharacterized genes<sup>2</sup>. (Image courtesy of Bruce W. Draper.)

Sönnichsen *et al.*<sup>2</sup> have addressed this limitation. They performed the herculean task of microinjecting dsRNAs corresponding to 98% of *C. elegans* genes into adult worms, and then scoring the progeny of microinjected animals for phenotypes ('progeny tests'). The injected worms' progeny were examined for embryonic lethality, larval arrest, morphological and neurological defects, and sterility. The authors thereby discovered the functions of 9% of worm genes. Compared to the feeding RNAi screens, microinjection identified a greater number of genes involved in embryonic development (1,266 genes in the injection study versus 929 in the feeding study). But fewer genes yielding larval and adult phenotypes were found. Thus, different RNAi methods may target different processes more or less effectively; other studies support this conclusion<sup>12–14</sup>.

The major advance provided by Sönnichsen *et al.* is a much more detailed analysis of all the early embryonic-lethal phenotypes reported. The embryos of *C. elegans* offer exceptional accessibility and clarity for observing early cell divisions, allowing analysis, at the cellular level, of the effects of gene knockouts. Thus, if an RNAi progeny test yielded an embryonic-lethal phenotype, Sönnichsen *et al.* proceeded to record the first two cell divisions using live cell imaging with standard light microscopy methods. The authors made over 40,000 time-lapse movies of developing embryos, following events in the first two cell divisions in each movie. These movies were then annotated for 45 distinct cellular phenotypes, including defects in such processes as passage through meiosis, centrosome attachment, mitotic spindle assembly and sister chromatid separation.

Obviously, these phenotypes go well beyond the designation of 'embryonic lethality' and point to specific cellular requirements. These data are available for searching and viewing on the authors' website<sup>15</sup>. Furthermore, the 661 genes important for early embryogenesis were divided into groups with related functions. Thus, gene pathways were shown to function in specific cell processes. For example, subunits of the proteasome, a complex involved in protein disposal, are highly enriched in the genes required for passage through meiosis, confirming that protein degradation is required during the meiotic cell cycle.

Sönnichsen and colleagues' data<sup>2</sup> offer remarkable insights into the cell biological functions of a large but manageable set of genes. This data set represents only a jumping-off point for even more specific functional studies, on a gene-by-gene basis. For example, determining the subcellular localization of the proteins required for a distinct process will provide a straightforward first step in determining whether such proteins

are directly responsible for the cellular event, or whether they regulate it from a distance.

However, the non-reproducibility of some phenotypes suggests that other gene requirements remain unidentified. It is also worth noting that RNAi-based phenotypes do not always recapitulate the 'null phenotype' that classical mutations can produce. Furthermore, the two genome-wide functional analyses in *C. elegans* have uncovered phenotypes for only 10% of the worm's genes<sup>2,10</sup>. This information will no doubt occupy many researchers for years to come — as will deciphering the function of the remaining 90% of the genes, and the rest of the DNA sequences in the genome, which await functional characterization by other methods. ■

Sean M. O'Rourke and Bruce Bowerman are at the

Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403-1229, USA.

e-mails: seanor@molbio.uoregon.edu;

bbowerman@molbio.uoregon.edu

1. Giaever, G. *et al. Nature* **418**, 387–391 (2002).
2. Sönnichsen, B. *et al. Nature* **434**, 462–469 (2005).
3. The *C. elegans* Sequencing Consortium. *Science* **282**, 2012–2018 (1998).
4. Lander, E. S. *et al. Nature* **409**, 860–921 (2001).
5. Kim, S. K. *et al. Science* **293**, 2087–2092 (2001).
6. Baugh, L. R., Hill, A. A., Slonim, D. K., Brown, E. L. & Hunter, C. P. *Development* **130**, 889–900 (2003).
7. Fire, A. *et al. Nature* **391**, 806–811 (1998).
8. Guo, S. & Kemphues, K. J. *Cell* **81**, 611–620 (1995).
9. Timmons, L. & Fire, A. *Nature* **395**, 854 (1998).
10. Kamath, R. S. *et al. Nature* **421**, 231–237 (2003).
11. Fraser, A. G. *et al. Nature* **408**, 325–330 (2000).
12. Gönczy, P. *et al. Nature* **408**, 331–336 (2000).
13. Simmer, F. *et al. PLoS Biol.* **1**, 77–84 (2003).
14. Maeda, I., Kohara, Y., Yamamoto, M. & Sugimoto, A. *Curr. Biol.* **11**, 171–176 (2001).
15. www.worm.mpi-cbg.de/phenobank2

## Earth science

# A different kind of foreshock

DelWayne R. Bohnenstiehl

Underwater sound recordings have been used to monitor transform faults in the equatorial Pacific, implicating a mechanism of foreshock generation distinct from that on most continental fault systems.

Nearly 40 years ago, a study of seismicity along oceanic transform faults provided telling evidence in support of the theory of plate tectonics<sup>1</sup>. Today, as exemplified by the work of McGuire and colleagues<sup>2</sup> (page 457 of this issue), seismologists are unravelling the mysteries of earthquake generation and testing the limits of prediction along these remote fault zones.

Earth's mid-ocean ridge system is where tectonic plates are moving apart and new sea floor is being created. Where one segment of this volcanic rift is offset from its neighbour, the differential motion of the plates is accommodated by a zone of lateral motion known as a transform fault (Fig. 1a). These fault zones are among the most conspicuous features on the sea floor, leaving scars that can often be traced across an ocean basin.

Like their continental cousins, such as the North Anatolian fault in Turkey and the San Andreas fault in the United States, most oceanic transforms can generate large earthquakes. Increasingly long-term seismic monitoring, however, has revealed dramatic differences between these two groups of faults. Most startlingly, an inventory of earthquakes along oceanic transforms documents a dramatic 'slip deficit', with about 85% of all transform motion occurring aseismically<sup>3</sup>. In contrast, for continental transforms, plate motion occurs almost exclusively during fault slippage events that generate earthquakes (Fig. 1b).

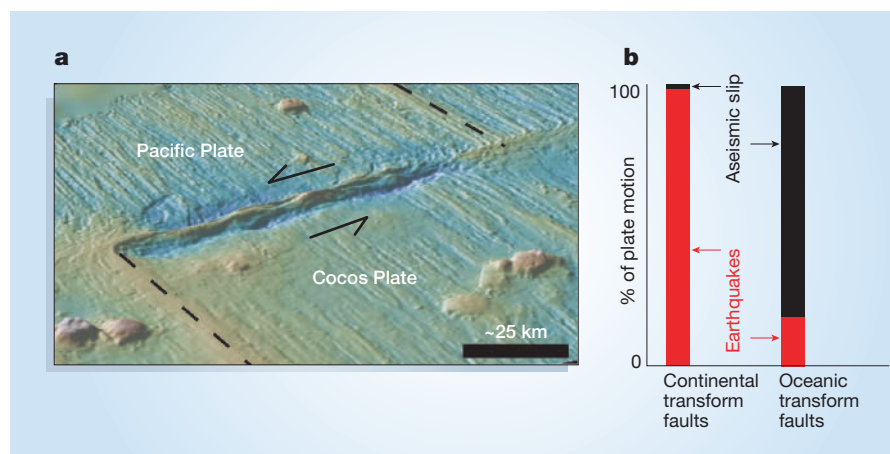
To examine this difference and explore its implications for understanding earthquake

initiation, McGuire *et al.*<sup>2</sup> turned to an array of moored hydrophones used by the US National Oceanic and Atmospheric Administration to monitor seismicity along parts of the equatorial East Pacific Rise<sup>4</sup>. These instruments take advantage of the more efficient propagation of sound in the oceans, relative to the solid Earth, to detect and locate earthquakes that would otherwise go unnoticed by distant land-based seismometers. Within the remote region of their study, the improvement in detection threshold provided by hydroacoustic monitoring is on the

order of two magnitude units. This enabled the authors to examine, for the first time, the spatial and temporal patterns of small-magnitude earthquakes (as little as 2.5 *m*) along a series of fast-sliding (more than 100 mm yr<sup>-1</sup>) oceanic transform faults.

An emerging view of seismicity is that there is no fundamental difference between foreshocks, mainshocks and aftershocks<sup>5,6</sup>. According to this perspective, any earthquake that occurs will subsequently produce other slip events, with the resulting number of 'triggered' earthquakes being a function of the size of the initial event. The magnitudes of the triggered events, however, are chosen randomly from the Gutenberg–Richter size–frequency distribution, which requires that  $\log N = a - bm$ , where  $N$  is the number of events of magnitude larger or equal to  $m$ ,  $b$  is an empirical constant often near 1, and  $a$  reflects the rate of local activity. Consequently, although smaller slip events should be triggered more frequently than larger ones, there is nothing to prevent a 'mainshock' from triggering an earthquake larger than itself, and the reclassification of this event as a foreshock is simply an afterthought. Unfortunately, according to this conceptual view, short-term prediction may be difficult, because it implies that foreshocks are not generated during a unique period of fault activity that foreshadows the impending mainshock<sup>7</sup>.

Although a single triggering mechanism can be applied within many continental fault systems<sup>5,6</sup>, McGuire and colleagues' analysis<sup>2</sup> suggests that foreshocks and aftershocks along oceanic transforms are generated by different processes. For example, they find that a mainshock along an oceanic transform produces fewer aftershocks than a comparably sized mainshock along a continental transform. Or, you might say, an oceanic aftershock sequence is 'depleted' relative to



**Figure 1 Finding fault.** a, An oceanic transform fault: perspective view showing bathymetry in the vicinity of the fast-slipping Clipperton Transform, near 10° N latitude on the East Pacific Rise. Oceanic transforms accommodate the differential motion of the tectonic plates (arrows) wherever mid-ocean-ridge spreading segments (dashed line) are significantly offset from one another. b, Histogram depicting the amount of plate motion accommodated by seismic (earthquake) and aseismic processes along typical continental and oceanic transforms.



one on land. If foreshocks and aftershocks are both generated by the same triggering process, we would expect to observe oceanic foreshock sequences that are similarly depleted. Instead, the ratio of foreshocks to aftershocks is roughly an order of magnitude larger than values observed within the continents. This observation is inconsistent with a single-mode triggering mechanism.

What could be producing the larger than expected number of foreshocks along oceanic transforms? One intriguing explanation is that they are triggered by a slow, creeping motion along part of the fault. These slow-rupture events are inefficient at producing seismic waves, and so are often termed quiet or silent earthquakes. This idea that aseismic fault movements could drive foreshock production is supported by the narrow time window (an hour or less) in which most foreshocks are observed and the tight spatial clustering of foreshocks in the vicinity of the mainshock<sup>2</sup>. In this view<sup>2,3</sup>, loud foreshock and mainshock earthquakes might be seen as aftershocks of quiet fault motions that were not detected by either the global seismic network or the regional hydrophone array.

This type of foreshock activity provides a way of tracking these quiet earthquakes along the transforms. Conceivably, it also opens the door to short-term prediction, because the transient slow-slip events represent a form of preparatory fault activity that precedes the mainshock. To test this idea, McGuire *et al.*<sup>2</sup> developed a simple prediction algorithm. An alert was declared each time the hydrophone array detected a transform earthquake (that is, a potential foreshock) larger than some threshold size, with a prediction that an event of large magnitude (a mainshock) would occur within a narrowly defined spatial and temporal window around the foreshock. This strategy proved surprisingly successful, with a predictive capability that outperformed more sophisticated algorithms implemented elsewhere.

Admittedly, these remote oceanic transform faults are far removed from population centres, limiting the immediate relevance of such predictions. Nonetheless, here we have a clear demonstration of the potential for short-term prediction for environments where foreshock activity may be driven by quiet slip events. Such settings could include several zones where tectonic plates are undergoing subduction<sup>8,9</sup>, which typically pose a more severe hazard because of their proximity to land.

DelWayne R. Bohnenstiehl is at the Lamont-Doherty Earth Observatory of Columbia University, PO Box 1000, Palisades, New York 10964, USA.  
e-mail: del@ldeo.columbia.edu

109, doi:10.1029/2004JB003110 (2004).

4. Fox, C. G., Matsumoto, H. & Lau, T.-K. A. *J. Geophys. Res.* **106**, 4183–4206 (2001).
5. Felzer, K. R., Abercrombie, R. E. & Ekstrom, G. *Bull. Seismol. Soc. Am.* **94**, 88–98 (2004).
6. Helmstetter, A., Sornette, D. & Grasso, J.-R. *J. Geophys. Res.*

108, doi:10.1029/2002JB001991 (2003).

7. Dodge, D. A., Beroza, C. G. & Ellsworth, W. L. *J. Geophys. Res.* **101**, 22371–22392 (1996).
8. Kanamori, H. & Cipar, J. *Phys. Earth Planet. Inter.* **9**, 128–136 (1974).
9. Ozawa, S. *et al. Science* **298**, 1009–1012 (2002).

## Physiology

# Do neural signals remodel bone?

Joel K. Elmquist and Gordon J. Stewler

The hormone leptin is best known for its influence on body weight. But it also controls bone mass, and recent work in mice is beginning to uncover the neuroendocrine systems involved.

The hypothalamus is the part of the brain that maintains homeostasis, which includes regulating the autonomic nervous system — the unconscious system for monitoring and controlling the state of the body's internal organs. A key input to the hypothalamus is leptin, a hormone that is produced in fat cells to signal energy sufficiency and which affects food intake, body weight and the neuroendocrine systems<sup>1–4</sup>. Previous work from Gerard Karsenty and colleagues showed that leptin also affects bone mass<sup>5,6</sup>: leptin-deficient *ob/ob* mice have high bone mass, and when leptin is replaced, their bone mass goes down. On page 514 of this issue (Eleftheriou *et al.*<sup>7</sup>), the team reports that leptin directly regulates bone physiology through control of the sympathetic nervous system, the arm of the autonomic nervous system best known for the fight-or-flight response.

The work is provocative for a number of reasons. First, the results establish that the sympathetic nervous system is required

for the effects of leptin on bone mass. The authors examined mice lacking the  $\beta_2$ -adrenergic receptor ( $\beta_2$ -AR) — one of several adrenergic receptors that respond to noradrenaline, a neurotransmitter released by sympathetic neurons. The mutant mice are resistant to the bone-reducing effects that leptin exerts via actions on the central nervous system (CNS). Second, the mice have increased bone mass at baseline, compared with normal mice, indicating that the sympathetic nervous system helps to set the level of bone mass, even under normal conditions.

Finally, and perhaps most strikingly, the mice are resistant to bone loss induced by the removal of their ovaries, and the consequent decrease of circulating oestrogen. Oestrogen deficiency in postmenopausal women causes rapid bone loss and predisposes them to osteoporotic fractures. Does menopausal bone loss require sympathetic nervous activity? Interestingly, ovariectomy in mice causes a dramatic loss of the nerves that usually penetrate bone<sup>8</sup>. These observations suggest

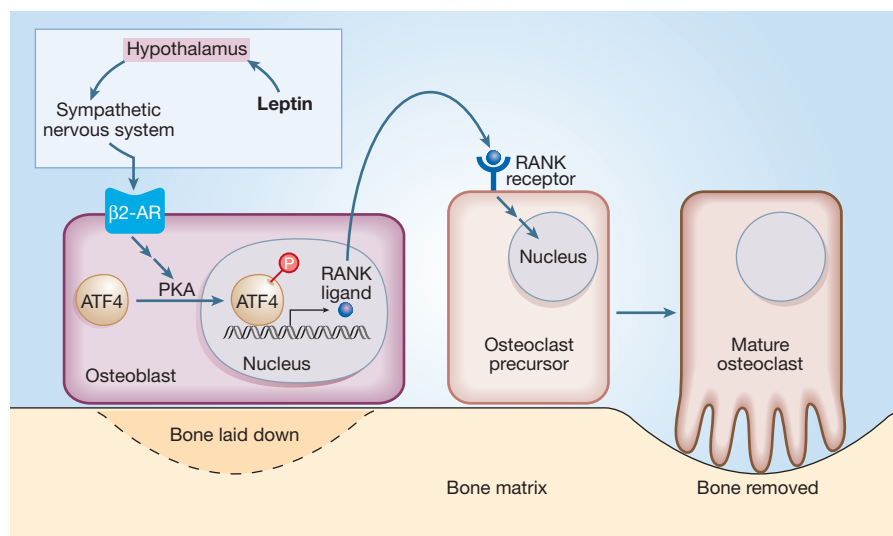


Figure 1 Brain and bone in sympathy. Eleftheriou *et al.*<sup>7</sup> find that leptin exerts its effects on bone through the sympathetic nervous system, which releases the neurotransmitter noradrenaline. Activation of one noradrenaline receptor, the  $\beta_2$ -adrenergic receptor ( $\beta_2$ -AR), causes osteoblasts to produce RANK ligand, through activation of protein kinase A (PKA) and phosphorylation (P) of the gene regulator ATF4. RANK ligand stimulates the formation of osteoclasts, which eat away the bone, so that overall bone mass diminishes.

1. Sykes, L. R. *J. Geophys. Res.* **72**, 2131–2153 (1967).

2. McGuire, J. J., Boettcher, M. S. & Jordan, T. H. *Nature* **434**, 457–461 (2005).

3. Boettcher, M. S. & Jordan, T. H. *J. Geophys. Res.*

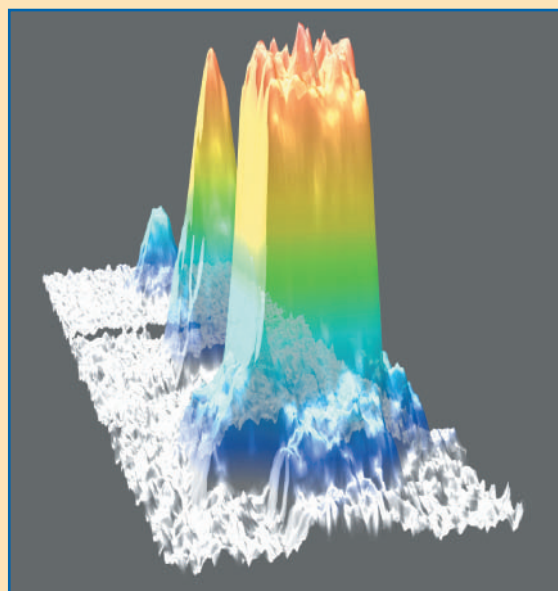
Condensed-matter physics

# Lab in a trap

A collection of fermions — sub-atomic particles with half-integer spin — will not all crowd into one energy state. Rather, two by two, they will populate the lowest-available states up to some cut-off energy, beyond which no ground-state fermions may stray. These populated states are bounded by the so-called Fermi surface. All fermionic systems have a Fermi surface, and an ultracold gas of potassium atoms trapped in an optical lattice is no exception, observe Michael Köhl and co-workers (*Phys. Rev. Lett.* **94**, 080403; 2005).

The lattice is composed of three standing waves of laser light held at right angles to each other. By controlling the depth of the lattice and the particle number, the underlying physics of quantum many-body systems — the interactions between atoms and their dynamics — can be investigated directly.

Köhl *et al.* show that they



can 'switch on' interactions between the particles in a Fermi gas. This results in a Fermi surface that, as more fermions are added to the trap, evolves from the smaller, rounded shape seen at the back of the image shown here to the larger, sharp-edged, square pillar in the foreground. Moreover, the interaction between two atoms in different spin states leads

to the higher-energy bands becoming populated, although the number of atoms involved is smaller than expected. This strongly interacting regime is not well understood, but Köhl and colleagues' results clearly demonstrate that optical lattices provide another approach to solving outstanding problems in condensed-matter physics. **May Chiao**

that oestrogen may be pivotal to the ability of the sympathetic nervous system to regulate bone formation. Moreover, if the sympathetic nervous system sets bone mass through  $\beta$ -adrenergic receptors, treatment with  $\beta$ -blockers (which are used to treat hypertension, for example) should improve bone mass. Whether this intriguing prediction is correct has not been settled by retrospective studies<sup>9–11</sup> and will require carefully designed clinical trials.

Bone continuously renews itself by a process called remodelling: old bone is eaten away (resorbed) by cells called osteoclasts, and new bone is laid down by cells called osteoblasts. It has been known for some time that the hypothalamus integrates inputs from the body and transmits signals that alter bone remodelling. This was previously thought to be mediated by altering the secretion of hormones, especially gonadal steroids. Eleftheriou *et al.*<sup>7</sup> report that the sympathetic nervous system is also involved, and they provide a molecular mechanism linking leptin, the sympathetic nervous system and decreased bone mass.

Hormones that control bone resorption often act indirectly, by stimulating osteoblasts to promote the formation of osteoclasts, rather than directly on osteoclasts

themselves. For example, Eleftheriou *et al.* report that adrenergic signals cause osteoblasts to secrete RANK ligand, the principal physiological inducer of osteoclast formation (Fig. 1). This effect requires protein kinase A, an enzyme that activates its target proteins by adding a phosphate group to them. In this case, the target is ATF4, a factor that is essential for osteoblast development and function. Once activated, ATF4 stimulates the production of RANK ligand.

Collectively, the results suggest that leptin induces bone loss via actions on the CNS and by regulating input from the sympathetic nervous system to bone. This is confirmed by the absence of leptin effects in mice lacking  $\beta$ 2-AR. However, leptin-deficient *ob/ob* mice have high rates of bone resorption (previously attributed to oestrogen deficiency), despite a reduction in  $\beta$ -adrenergic signalling that should lead to decreased resorption.

A final interesting component of the work by Eleftheriou and colleagues may help to explain this apparent paradox. They report that mice lacking the leptin-regulated neuropeptide CART (for 'cocaine- and amphetamine-regulated transcript'), have a low bone mass at baseline. Because levels of the RNA encoding CART are very low in leptin-deficient mice, one might predict that CART

mediates the effects of leptin. By contrast, the authors find that mice without CART have increased bone resorption in response to leptin. These observations suggest that neurons expressing CART act to inhibit bone loss, including that mediated by leptin.

The CART neurons involved in bone loss have yet to be identified. One obvious target is the CART neurons in a region of the hypothalamus called the arcuate nucleus<sup>12</sup>. These neurons are stimulated by leptin and also express POMC, the prototypic CNS target of leptin. Interestingly, these neurons directly target sympathetic preganglionic neurons (the first neurons in the autonomic neuron chain)<sup>13</sup>. However, it is unclear whether they mediate leptin's effects on bone, as mice that lack leptin receptors only in POMC/CART neurons do not seem to have bone abnormalities<sup>14</sup>.

CART expression is widespread in the brain and spinal cord, including sympathetic preganglionic neurons themselves, and in several other regions of the hypothalamus including the 'ventral premammillary nucleus', which has neurons that express high levels of receptors for sex steroids<sup>15</sup>. Given the proposed interaction of oestrogen and the autonomic nervous system, it is intriguing to speculate that hypothalamic cell groups (including CART neurons) might integrate a number of signals, including leptin and oestrogen, to coordinate the CNS control of bone mass through its output to sympathetic preganglionic neurons.

Several points remain to be reconciled with this model of the regulation of bone mass. For example, obese women have high leptin levels and high bone mass; lean women with very low leptin levels (such as marathon runners) have low bone mass. In addition, whether leptin has direct effects on bone cells remains to be settled<sup>5,16</sup>. Nonetheless, Eleftheriou and colleagues' findings provide support for the notion that the hypothalamus and its control of the autonomic nervous system are key to the regulation of bone homeostasis. ■

Joel K. Elmquist and Gordon J. Stewler are at the Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts 02215, USA. e-mail: jelmquis@bidmc.harvard.edu

1. Ahima, R. S., Saper, C. B., Flier, J. S. & Elmquist, J. K. *Front. Neuroendocrinol.* **21**, 263–307 (2000).
2. Friedman, J. M. *Nutr. Rev.* **60**, S1–S14 (2002).
3. Flier, J. S. *Cell* **116**, 337–350 (2004).
4. Zigman, J. M. & Elmquist, J. K. *Endocrinology* **144**, 3749–3756 (2003).
5. Takeda, S. *et al. Cell* **111**, 305–317 (2002).
6. Ducey, P. *et al. Cell* **100**, 197–207 (2000).
7. Eleftheriou, F. *et al. Nature* **434**, 514–520 (2005).
8. Burt-Pichat, B. *et al. Endocrinology* **146**, 503–510 (2005).
9. Pasco, J. A. *et al. J. Bone Miner. Res.* **19**, 19–24 (2004).
10. Rejnmark, L. *et al. Calcif. Tissue Int.* **75**, 365–372 (2004).
11. Schlienger, R. G., Kraenzlin, M. E., Jick, S. S. & Meier, C. R. *J. Am. Med. Assoc.* **292**, 1326–1332 (2004).
12. Kristensen, P. *et al. Nature* **393**, 72–76 (1998).
13. Elias, C. F. *et al. Neuron* **21**, 1375–1385 (1998).
14. Balthasar, N. *et al. Neuron* **42**, 983–991 (2004).
15. Elias, C. F. *et al. J. Comp. Neurol.* **432**, 1–19 (2001).
16. Reseland, J. E. & Girdeladze, J. O. *FEBS Lett.* **528**, 40–42 (2002).



## Biodiversity

## Gut feeling for yeasts

Teun Boekhout

The startling news that has emerged from studies of the intestines of beetles is a reminder of how little is known about the diversity of even such comparatively well-characterized groups as the yeasts.

There is no doubt that many species of fungi have yet to be described, yeast species included. Many of them must lurk in habitats and hosts that have still to be investigated. But who would have thought to look in the guts of beetles? Suh and colleagues did so, as they report in *Mycological Research*<sup>1</sup>, and have found a hotspot of yeast diversity. About 1,000 yeast species are known worldwide, to which Suh *et al.* add at least 200 more, identified from a variety of beetles sampled in the southeastern United States and Panama. Many of the yeasts belong to such familiar genera as *Candida* and *Pichia*, some species of which are respectively human pathogens and agents used for the production of drugs.

Yeasts are among the best-studied groups of microbes, but this research provides a dramatic indication of how limited our knowledge is. Apart from their traditional importance in wine, bread and beer making, they are also used in many (agro)industrial and medicinal applications, and as workhorses in biological research. Suh and colleagues' studies on beetles show that efficient sampling of poorly explored habitats or geographical areas can yield an extraordinary extension to our knowledge of microbial biodiversity.

Since 1820, about 3,000 yeast species have been described (Fig. 1), many of which were later considered to be identical to previously described species, and consequently regarded as synonyms. Two recent monographs on yeasts accept only approximately 700 species<sup>2,3</sup>, but this figure has since risen closer to 1,000. Our present knowledge is heavily biased, however. The great majority of known species come from Western Europe, Japan and North America. Of the strains maintained in the collection of the Centraalbureau voor Schimmelcultures (CBS), which is the result of a century of work, 45% originated from clinical sources or from products such as wine, fermented foods, bread, fruit juices and dairy products, with 55% from natural substrates — plants, soil, fruits, animals and water<sup>4</sup>. Insect sources accounted for only 6% of the isolates in the CBS yeast collection. Clearly, this is only the tip of a very large iceberg: Suh *et al.*<sup>1</sup> consider that further sampling of beetles alone could result in the identification of another 350–500 yeast species.

In recent times, most researchers interested in yeast biodiversity have made effective

use of DNA-based taxonomic concepts<sup>5</sup>. Since the 1990s, a huge effort has been made with all accepted yeast species to sequence the specific regions of DNA that encode a cell component known as ribosomal RNA — these are standard sequences in systematics. The result is that comprehensive rDNA databases now exist for yeasts<sup>6,7</sup>. These molecular studies have produced an unparalleled increase in the number of yeast species described in the years 2000–05 (Fig. 1), even before we had news of the high biodiversity observed in beetle guts.

That news adds an extra dimension to our appreciation of yeast ecology. Suh *et al.* found that some clades of yeasts contain many insect-related species (a clade being a grouping consisting of an ancestor and all its descendants). Examples are the *Pichia guilliermondii*, *P. stipitis*, *Candida tanzawaensis* and *Stephanoascus-Arxula-Zygoascus* clades in the fungal phylum Ascomycota, and the *Trichosporon* clade within the phylum Basidiomycota. The *C. tanzawaensis* clade has expanded from a single strain of just one known species to 164 isolates representing

39 possible species. One wonders whether some of the insect-related species were described in the past, but traditional means of classification erroneously interpreted them as being identical to other known species. This may be true for some of the species in, for example, the *P. guilliermondii* clade, and it will be necessary to sequence the 'type' (reference) strains of all species considered to be synonyms.

Given the continuing development of molecular tools for the rapid identification of microbes, studies of yeast diversity in habitats and regions previously unexplored, or only scantily so, have become more feasible and will pay rich dividends. Large parts of Africa, Antarctica, Asia, Australia and Latin America are mainly virgin territory in this respect<sup>8</sup> — it could be that only 1% of yeast species have been described<sup>9</sup>. This endeavour will not just be for the education of biologists concentrating on biodiversity, but will also enhance our knowledge of ecological interactions — as, for instance, is the case in research<sup>10</sup> on xylose-fermenting yeasts in beetle guts, which may be involved in the breakdown of hemicellulose, a component of plants, and thus contribute to the decomposition of plant biomass. Moreover, some as-yet-undiscovered species are likely to be sources of useful compounds or as agents for biological control.

As far as the work of Suh *et al.*<sup>1</sup> is concerned, extra relevance stems from the fact that insects are also major vectors for the dispersal of yeasts, among them those involved

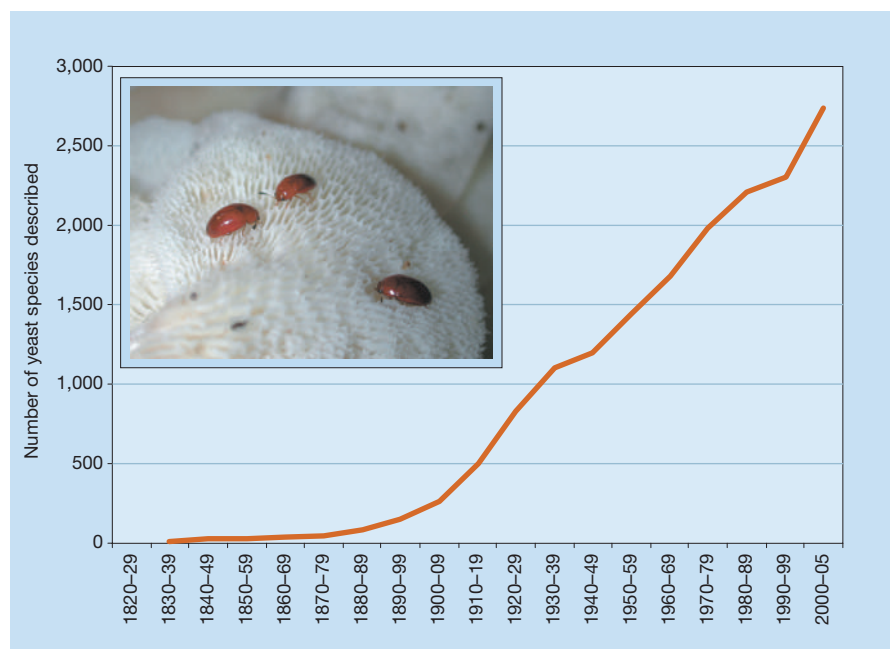


Figure 1 Yeast on the rise — the cumulative increase in yeast species described since 1820 (the total includes synonyms). Since 2000, application of DNA-based methods has produced a rapid increase in the yeast species recognized; the last time span covers only five years rather than ten, and includes the new identifications from the hyperdiverse guts of beetles. Blips in the increase in descriptions occurred during the Second World War and at the end of the twentieth century. Inset, beetles of the genus *Mycotretus*, family Erotylidae, one of the taxonomic groups of beetle sampled by Suh and colleagues<sup>1</sup>.

in fermentation and food spoilage. So a better understanding of yeast biodiversity in insects should help in controlling these processes. All in all, the new observations are a clear signpost to a highly promising route in the quest to identify hidden microbial biodiversity. ■

Teun Boekhout is in the Centraalbureau voor Schimmelcultures, Uppsalalaan 8, 3584 CT Utrecht, and the Division of Acute Medicine and Infectious Diseases, Utrecht University, The Netherlands. e-mail: boekhout@cbs.knaw.nl

1. Suh, S.-O., McHugh, J. V., Pollock, D. D. & Blackwell, M. *Mycol. Res.* **109**, 261–265 (2005).

2. Kurtzman, C. P. & Fell, J. W. *The Yeasts: A Taxonomic Study* 4th edn (Elsevier, Amsterdam, 1998).
3. Barnett, J. A., Payne, R. W. & Yarrow, D. *Yeasts: Characteristics and Identification* (Cambridge Univ. Press, 2000).
4. CBS databases [www.cbs.knaw.nl](http://www.cbs.knaw.nl)
5. Blazer, M. L. *Phil. Trans. R. Soc. Lond. B* **359**, 669–679 (2004).
6. Kurtzman, C. P. & Robnett, C. J. *Antonie van Leeuwenhoek* **73**, 331–371 (1998).
7. Scorzetti, G., Fell, J. W., Fonseca, A. & Statzell-Tallman, A. *FEMS Yeast Res.* **2**, 495–517 (2000).
8. Hawksworth, D. L. *Stud. Mycol.* **50**, 9–18 (2004).
9. Fell, J. W., Boekhout, T., Fonseca, A. & Sampaio, J. P. in *The Mycota VII*, Part B, 2nd edn (eds McLaughlin, D. J., McLaughlin, E. G. & Lemke, P. A.) 3–35 (*Systematics and Evolution*, Part B) (Springer, Berlin, 2001).
10. Suh, S.-O., Marshall, C. J., McHugh, J. V. & Blackwell, M. *Mol. Ecol.* **12**, 3137–3146 (2003).

## Nanotechnology

# New spin on correlated electrons

Ronald M. Potok and David Goldhaber-Gordon

In the Kondo effect, the flow of electrons in a solid is modulated by magnetic impurities. Nanostructures such as carbon nanotubes can be designed to obtain even more complex versions of this intriguing effect.

The behaviour of electrons is difficult to predict. Undergraduates routinely calculate the quantum states of a hydrogen atom, but even a helium atom with two electrons displays complex dynamics. How much more complex are solids, with electrons in the billions. Nanotechnology now enables researchers to create controlled, or

tunable, models of electronic behaviour in solids. On page 484 of this issue, Jarillo-Herrero *et al.*<sup>1</sup> present perhaps the most sophisticated example of this approach to date, studying the flow of electrons through a carbon nanotube that is made to act like an exotic magnetic atom.

In attempting to explain the electronic properties of materials, theorists often construct ‘microscopic’ models, comprising electrons that can move freely (become delocalized), sit on localized sites and make transitions between these states. Such models offer powerful insights but also have drawbacks when applied to conventional bulk materials. In particular, electrons on many sites can interact with each other, making complete calculations impractical; and parameters such as an electron’s tunnelling rate and interaction strength are not tunable, and can be difficult to measure precisely. Nanotechnology can remedy these limitations: by building submicrometre-sized structures in which one or a few electrons are isolated, researchers can know everything about the relevant electronic states, and can study electrons at a single site. Perhaps most importantly, nearly all of the significant parameters can be designed, measured and often tuned *in situ*.

The interaction of mobile electrons with magnetic impurities — atoms or ions with a non-zero magnetic moment — in a host metal has been a theme in solid-state physics for decades. In 1961, Anderson proposed that a magnetic impurity could be modelled as an electronic quantum state localized at a single site within a crystal. According to the Pauli exclusion principle, the site can be occupied by up to two electrons (with opposite spins), but Coulomb repulsion can make

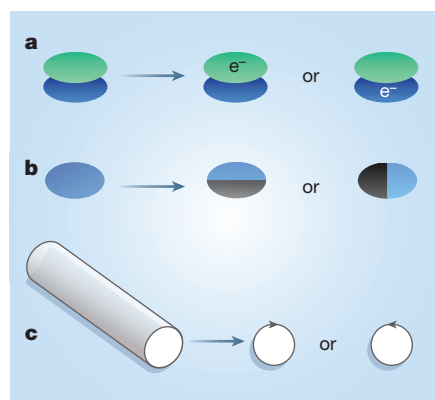


Figure 1 Degeneracy and the Kondo effect. The basis of the Kondo effect is the screening of a localized twofold degeneracy by surrounding delocalized electrons. The degeneracy normally comes from spin, but with the advent of nanotechnology, researchers can create and study other types of degeneracy; a–c give examples. a, An electron can reside on either of two sites, but not both simultaneously because of Coulomb repulsion<sup>8</sup>. b, c, A spatial symmetry produces degenerate orbital levels<sup>11</sup>. b, The symmetry is the rotational symmetry of a circular quantum dot<sup>9</sup>. This is analogous to the rotational symmetry of a hydrogen atom that produces degenerate *p* orbitals. c, The lattice symmetry of a nanotube gives rise to a twofold orbital degeneracy<sup>7</sup>. Non-spin degeneracy may also give rise to the Kondo effect in bulk systems<sup>14</sup>.



## 100 YEARS AGO

*Die Kalahari.* Now if we could imagine that Mr. Shandy the elder were alive, this is a book that, like many another of its class, would have delighted him. Hereby he could have proved triumphantly to Yorick the potency of that great scheme of education — that “north-west passage to the intellectual world” — which he propounded so enthusiastically upon a memorable occasion. His scheme, it will be remembered, was that upon every substantive in the dictionary (so gravely misunderstood by the Corporal and Uncle Toby) should be brought to bear exhaustively:— “Every word, Yorick, by this means, you see, is converted into a thesis or an hypothesis:— every thesis and hypothesis have an offspring of propositions;— and each proposition has its own consequences and conclusions; every one of which leads the mind on again, into fresh tracts of enquiries and doubtings.— ‘The force of this engine,’ added my father, ‘is incredible...’ ”

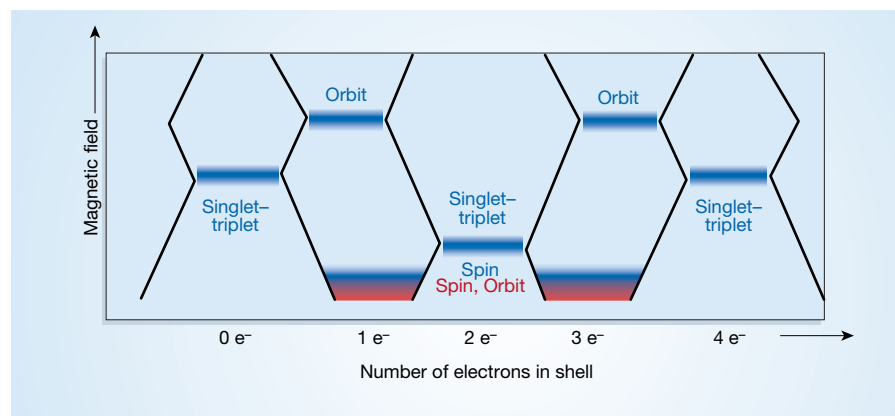
From *Nature* 23 March 1905.

## 50 YEARS AGO

Dr. J. R. Oppenheimer’s address was ostensibly on... “Man’s Right to Knowledge and the Free Use Thereof”; but he dealt more specifically with the prospect in the arts and sciences, and above all he emphasized the isolation of the specialist from the broad path of human knowledge and intercourse. With more means of communication than ever before, we can speak to each other less easily and less effectively. While more people are engaged in scientific research to-day... the frontiers of science are now separated by long years of study, by specialized vocabularies, arts, techniques and knowledge from the common heritage even of the most civilized society... The difficulties which face us are thus derived from the growth of skill, of power, and of our understanding of that world, if not of the conditions which govern their exercise. To assail such changes is futile: we need to recognize them and to take account of our resources; nor does Dr. Oppenheimer appear to put high among such resources the mass media of radio and television, the cinema, best-seller or the world tours of successful theatrical productions. Although art and science may be purveyed by such means, they have also an inhumanizing influence and can easily destroy the individual sense of beauty.

From *Nature* 26 March 1955.





**Figure 2** Many kinds of Kondo. For any number of electrons in a shell of a nanotube, the low-energy states can be identified. At zero magnetic field, a shell with either one or three electrons displays a fourfold degeneracy, giving rise to a novel fourfold Kondo effect (red). Applying a magnetic field first breaks orbital degeneracy, reducing fourfold Kondo to conventional twofold spin Kondo (blue), and then destroys even the twofold spin Kondo. Higher magnetic fields can bring together electronic states that were not degenerate at zero magnetic field, producing a variety of Kondo effects for both even and odd electron occupancy. Although all these Kondo effects are twofold (blue), none emerges from conventional spin degeneracy. Their various origins are explained in the Supplementary Information to ref. 1.

it energetically favourable for only one electron with arbitrary spin to reside there. If its spin opposes that of a nearby delocalized electron, the localized electron can lower its energy by tunnelling off the site and back on again. At low temperature, the electron ensures that it can tunnel by pairing with a single delocalized electron of opposite spin to form a state of zero total spin (a 'singlet') — a phenomenon known as the Kondo effect. In Anderson's microscopic model, the stability of the singlet ground state (and hence its temperature-dependent effect on electrical resistance) can be calculated in terms of basic parameters such as the binding energy of the localized electron and its rate of tunnelling between localized and delocalized states<sup>2</sup>.

Now, however, experimental physicists can design and fabricate single artificial magnetic impurities, such as semiconductor quantum dots<sup>3</sup>, molecules<sup>4,5</sup> and nanotubes<sup>6</sup>, with substantial control over the basic parameters of Anderson's model. In each of these systems, conducting leads attached to the artificial impurity play the same role as the host metal of a traditional magnetic impurity. Electron flow from one lead to the other through the artificial impurity serves as a powerful probe of local electronic states. The tunability of such nanostructures, and their amenability to electrical measurement, have allowed highly quantitative studies of the Kondo effect<sup>7</sup>, and have spurred the construction of new Kondo systems not described by the minimal Anderson or Kondo model (see the Supplementary Information to ref. 1 for a review).

The essence of the Anderson model is a localized state with a twofold degeneracy (two different quantum states of equal energy), coupled to a sea of electrons with the

same degeneracy. Normally this degeneracy comes from spin, but recently it has been verified that other twofold degeneracies can support the Kondo effect without spin<sup>8</sup>. This work forms the foundation for the achievements of Jarillo-Herrero *et al.*<sup>1</sup>.

The first non-spin Kondo implementation to be devised linked two sites capacitively, so that a charge could exist on one or the other, but not on both simultaneously<sup>8</sup> (Fig. 1a). Since then, Sasaki *et al.*<sup>9</sup> with a quantum dot, and now Jarillo-Herrero *et al.*<sup>1</sup> with a nanotube, have used a single site at which two degenerate orbital states act as the localized 'magnetic impurity'. Here the degeneracy stems from a spatial symmetry in the device, either associated with the structure of the material (Fig. 1c) or created by nanoscale patterning (Fig. 1b). In these experiments, a Kondo effect can occur because the local degree of freedom is also reflected in the nearby leads. With a traditional magnetic impurity, if the spin of the localized electron is pointed up, a nearby, delocalized electron will direct its spin down in response. Similarly, if there is an extra charge on the top site of Figure 1a, an electron in the nearby leads will move towards the bottom.

Orbital degeneracy does not preclude spin degeneracy. The same surrounding delocalized electrons that electrostatically screen the local orbital state also magnetically screen the localized spin, so that the two degrees of freedom become intimately connected, together forming a new fourfold degenerate Kondo effect<sup>10,11</sup>. Jarillo-Herrero *et al.* have demonstrated that the spin and orbital states of a carbon nanotube can be finely controlled so as to observe this effect.

In an ideal nanotube, sets of four electronic states can be grouped together into a shell, much like those of an atom. All four

states are degenerate, with two choices for spin and two for orbital state. Electrons can be added one at a time to the shell by fine control of the voltage on a nearby electrode. The first electron can reside in any of four states, with strong coupling to the metal leads producing a fourfold Kondo system. If a second electron is added to the nanotube, the Kondo effect vanishes, suggesting that the two electrons form a unique, non-degenerate ground state. The fourfold Kondo effect returns with the addition of a third electron, because now any one of the four states can be left empty. A fourth electron completes the shell, removing all degeneracy and thus eliminating the Kondo effect.

In experiments, the Kondo effect is generally identified by a distinctive logarithmic temperature dependence of electrical resistance. Fourfold Kondo is predicted (and now observed) to have the same temperature dependence as conventional twofold Kondo, but the two systems can be distinguished by their very different responses to a magnetic field. Applying a magnetic field creates and breaks a variety of degeneracies (Fig. 2), allowing Jarillo-Herrero *et al.* to observe not only fourfold Kondo but also three other types of Kondo effect associated with different degeneracies in a single nanotube.

Researchers can unambiguously identify the states of several electrons in nanostructures, and even manipulate those states. Electrons in nanostructures can then be mapped onto simple microscopic models, which successfully predict a variety of novel correlated-electron behaviour. The coming years should see researchers building more complex models, and addressing phenomena that are at the centre of modern condensed-matter theory, such as quantum phase transitions<sup>12,13</sup> and decoherence. ■

Ronald M. Potok is in the Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA.

e-mail: potok@stanford.edu

David Goldhaber-Gordon is in the Department of Physics, Stanford University, Stanford, California 94305, USA.

e-mail: goldhaber-gordon@stanford.edu

- Jarillo-Herrero, P. *et al.* *Nature* **434**, 484–488 (2005).
- Haldane, F. D. M. *Phys. Rev. Lett.* **40**, 416–419 (1978).
- Goldhaber-Gordon, D. *et al.* *Nature* **391**, 156–159 (1998).
- Liang, W. *et al.* *Nature* **417**, 725–729 (2002).
- Park, J. *et al.* *Nature* **417**, 722–725 (2002).
- Nygard, J., Cobden, D. H. & Lindelof, P. E. *Nature* **408**, 1342–1346 (2000).
- Goldhaber-Gordon, D. *et al.* *Phys. Rev. Lett.* **81**, 5225–5228 (1998).
- Wilhelm, U., Schmid, J., Weis, J. & von Klitzing, K. *Physica E* **14**, 385–390 (2002).
- Sasaki, S. *et al.* *Phys. Rev. Lett.* **93**, 017205 (2004).
- Li, Y. Q., Ma, M., Shi, D. N. & Zhang F. C. *Phys. Rev. Lett.* **81**, 3527–3530 (1998).
- Zarand, G., Brataas, A. & Goldhaber-Gordon, D. *Solid State Commun.* **126**, 463 (2003).
- Jeong, H., Chang, A. M. & Melloch, M. R. *Science* **293**, 2221–2223 (2001).
- Craig, N. J. *et al.* *Science* **304**, 565–567 (2004).
- Matsushita, Y., Bluhm, H., Geballe, T. H. & Fisher, I. R. preprint at <http://arXiv.org/cond-mat/0409174>.

## Obituary

## Hubert Curien (1924–2005)

“M. Hubert Curien is an unassuming, almost shy, man, with the qualities of a terrier” — that was one summary of the personality of the French crystallographer who was among the main architects of both French and European science (see *Nature* 346, 126–127; 1990). Curien, who died on 6 February at the age of 80, was one of a rare breed. He was a scientist with an undisputed flair for politics and diplomacy at the highest level, heading up CERN, the European particle physics facility, and the European Space Agency (ESA). For much of a decade, he was also one of the most effective science ministers France has had, enjoying the ear of president François Mitterrand.

But much of Curien's success also resided in his disarming humility, and his wry sense of humour, simplicity and *savoir vivre*. As a journalist, I was always struck by the fact that, despite the burdens of office, Curien always seemed to have time to discuss and to listen. He was deeply appreciated by the research community, who have fond memories of his terms of office as science minister under four socialist prime ministers — Laurent Fabius (1984–86), Michel Rocard (1988–91), Edith Cresson (1991–92) and Pierre Bérégovoy (1992–93).

His mix of talents allowed Curien slowly, but surely, to take his visions forward, not least his ambitions for European space science and exploration, in the complex context of European politics — where, to this day, national interests and bureaucracy stymie grand ideas. Back in the 1970s, who would have imagined that Ariane, Europe's fledgling rocket launched from ESA's spaceport in Kourou, French Guiana, would grow to dominate the world market and give Europe independent access to space?

Curien was born in Cornimont in the Vosges region of eastern France in 1924; his mother was a headmistress and his father a tax collector. In 1944, he enlisted in the French resistance. The Second World War left a lasting impression on Curien, forging his conviction of the need to construct a European Union.

Starting in 1945, Curien read physics at the Ecole Normale Supérieure in Paris, after which he went on to pursue a research career in crystallography (discovering three crystalline forms of gallium, and having a mineral, curienite, named after him). Throughout his subsequent long political career, he continued to work in his laboratory at the University of Pierre

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

### Father of the Ariane rocket programme, and champion of a Europe united in science

and Marie Curie in Paris. Attached to teaching, Curien also maintained his courses at the university until 1994.

His talent as a research administrator appeared early, and he quickly rose through the ranks of the CNRS, France's national research agency, becoming its director-general in 1969 and initiating a modernization of the agency. He was one of the founders of the European Science Foundation, based in Strasbourg, and was its first chairman from 1979 to 1984, and from 1994 to 1996 he was president of the board of CERN. Janez Potocnik, the European Union research commissioner, paid tribute to Curien's efforts: “If something like a scientific Europe exists today, it is in great part thanks to him. He has played a key role in the history of European scientific cooperation.”

Curien's destiny as an architect of European space policy was to be shaped by his time as head of France's space agency, the CNES, during 1976–84, and his overlapping role as the first chairman of the board of ESA from 1981 to 1984. Curien was concerned that the United States and Russia would dominate the space industry, and believed that Europe had to have its own independent capacity to launch satellites. He and Roy Gibson, ESA's first director-general, persuaded other countries to join ESA by creating the system of ‘just returns’ for industrial contracts. He oversaw the first launch of Ariane, and decided to create Arianespace.

Curien also played a critical role in the creation of ESA's flagship space science

programme at a ministerial conference in Rome in 1985. Roger Bonnet, who headed ESA's science division from 1983 to 2001, remembers how, with the conference deadlocked, Curien came back from lunching with ministers opposed to the costs of the new programme with a deal giving it a massive 5% annual increase in budget for the next ten years. “Curien's diplomatic skills had a big effect on the birth of European space science,” says Bonnet.

At the same time, Curien launched France into Earth observation with the SPOT series of satellites, and oversaw the setting up of the European Astronaut Corps. Less successful were his dreams for a European space shuttle, Hermes. With the costs of the craft soaring, Curien, in his second spell as science minister from 1988 to 1993, did not hesitate to cancel what had been his own brainchild. Curien oversaw many of the heady days of Europe's fledgling space programme. Following the launch of the first Ariane above the jungle in Guiana in 1979, he and his colleagues celebrated with a tropical snowball fight, using the liquid hydrogen and oxygen dispensed from Ariane's fuel tanks.

It was after another successful rocket launch in Kourou that Curien caught the eye of Laurent Fabius. Curien was later to remark, “I climbed onto a barrel to say a few words as was the custom, and was welcomed by thunderous applause... Fabius, who had to create a new government, said that if all it took was for me to climb onto a barrel to be acclaimed, I would be a very popular minister.”

Curien oversaw a massive investment in French research by Mitterrand, with spending on science rising from 1.97% of gross domestic product in 1981 to 2.42% in 1992. This and a raft of reforms reorganizing the research agencies were enshrined in a 1982 law on science, which made research a national priority and underpinned a resurgence of French science in the 1980s. Curien was appreciated across the political spectrum, and for example was nominated as chair of the national scientific council of defence in 1986 by the neo-Gaullist government of Jacques Chirac, while at the same time being chairman of a campaign for the re-election of Mitterrand in the poll of 1988.

Following Curien's death, Jean-Pierre Raffarin, Chirac's current prime minister, spoke for both the political and scientific communities in France: Curien was, he said, “a man of conviction who knew how to overcome all divisions in the national interest”.

Declan Butler

Declan Butler is a senior reporter for *Nature* based in Paris.



## Networks

### More than the sum of its parts

*Proc. Natl Acad. Sci. USA* **102**, 4221–4224 (2005)

Claims of scale-free network behaviour need to be approached with caution, say Michael P. H. Stumpf and colleagues.

Many networks in nature and in technology grow by the addition of links to nodes that are already highly connected. Popular websites, for example, are more likely to attract links as new pages are added to the World Wide Web. This 'preferential attachment' process leads to networks that are 'scale-free': there is no preferred scale to the average connectivity of nodes, which instead have a probability distribution of connections described by a power law.

Many networks have been suggested to have this structure, including transport systems, food webs and the networks formed by interactions between proteins or genes. But rarely is such a scale-free network topology identified by looking at the entire network; often, for example, protein networks are constructed from just 10–20% of the total protein complement of an organism.

Stumpf *et al.* point out that, for scale-free networks, such incomplete sampling can be misleading, because subsets of such networks are not themselves scale-free. This contrasts with the properties of other classes of network, such as random graphs, where subsets do reflect the structure of the whole.

Philip Ball

## Systems biology

### Many paths, one destination

*Phys. Rev. Lett.* (in the press)

How does a single type of progenitor cell form the diverse array of cells that make up an organism? Sui Huang *et al.* have approached this question by tracking the route taken by cultured cells differentiating from one cell type into another.

When HL60 cells (a human cell line) are treated with either the solvent dimethyl sulphoxide or the hormone all-*trans* retinoic acid they are stimulated to become neutrophils (a kind of white blood cell). Huang *et al.* simultaneously monitored the expression of 12,600 genes in cells undergoing these two treatments. Some 3,841 genes showed significant changes in expression and these were used as the axes on which the cells' journey from HL60 cells to neutrophils were plotted.

The expression profiles of the neutrophils produced by either stimulant were the same. But this shared end point was reached by very different routes. Despite similar beginnings, the expression paths plotted for each population quickly diverged, ultimately

## Biochemistry

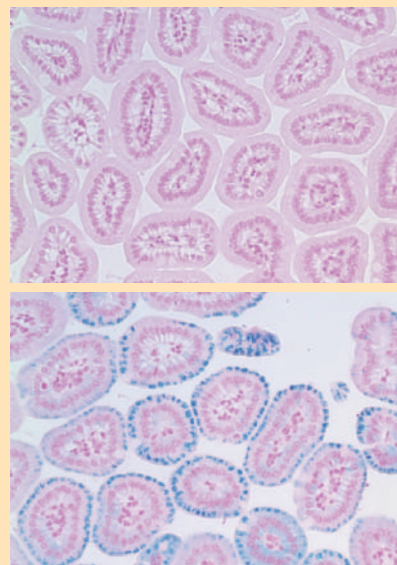
### No escape for iron

*Cell Metab.* **1**, 191–200 (2005)

A single protein — ferroportin — has been identified as the major exporter of iron from cells, and so is implicated in the accumulation of iron in people suffering from the hereditary disease haemochromatosis. If left untreated, iron overload can cause serious organ damage and is sometimes fatal. The liver and kidneys cannot eliminate iron from the body, as they do other metals, and so cells throughout the body must be able to store and release it when appropriate.

Adriana Donovan and colleagues investigated the role of ferroportin using genetically designed mice with the protein switched off. The animals died early in fetal development because iron was not transferred from mother to embryo. Using a genetic trick, Donovan *et al.* managed to produce live-born mice lacking ferroportin; these animals had large amounts of iron in their intestinal cells (the blue stain on the bottom image) compared with intestinal cells from normal mice (top).

Inactivating ferroportin only in the intestine caused mice to develop anaemia. This result is notable because the intestine normally absorbs



iron for transport to the bone marrow, where haemoglobin is made; the intestinal cells could clearly absorb iron, but not release it. The authors say that these mouse models provide the first *in vivo* evidence that the iron export function of ferroportin is essential for mammalian iron homeostasis and survival.

Roxanne Khamis

**reconverging on their common goal from diametrically opposite 'directions'. Thus, cell differentiation seems to be less the observance of a carefully prescribed path, and more like a ball that can take multiple paths as it rolls over a hill from one valley, or stable cell type, to the next.** Christopher Surridge

## Condensed-matter physics

### Chromium condensate

Preprint at <http://arXiv.org/abs/cond-mat/0503044>

When gases of certain atoms are cooled to extremely low temperatures, they can form a Bose–Einstein condensate, in which all the atoms share the same quantum state. These condensates have unusual properties, such as superfluidity (flowing without friction) and superconductivity (conducting electricity without resistance). Axel Griesmaier *et al.* have now created a condensate using chromium atoms.

Bose–Einstein condensates are usually prepared from alkali metals such as rubidium and potassium. These atoms each have a single unpaired electron, giving them very weak magnetic properties. Griesmaier and colleagues' condensate consists of more than 50,000 chromium atoms at a temperature of about 700 nanokelvin. Each atom carries six unpaired electrons, making their magnetic dipole–dipole interactions 36 times as strong as those of the alkali metals.

This phenomenon should allow researchers to probe the magnetic properties of the condensate, Griesmaier *et al.* suggest. They also expect to be able to adjust both

the short- and long-range interactions of the condensate using magnetic fields. And they say that improved lithography processes could result from creating a 'laser' of identical chromium atoms, which could be used to position single atoms exactly on a surface.

Mark Peplow

## Protein dynamics

### Folding free of charge

*J. Am. Chem. Soc.* doi:10.1021/ja043804d (2005)

Many proteins have surfaces peppered with electrically charged groups. It is tempting to imagine that, by offering favourable interactions with water, these help to stabilize the folded state and to promote solubility. Oppositely charged groups might also provide a kind of glue by attracting each other, whereas a protein covered with like charges might be destabilized by electrostatic repulsions.

Katherine L. Gudiksen *et al.* challenge all of these notions. They show that converting 18 of the 27 lysine-NH<sub>3</sub><sup>+</sup> groups on the surface of the enzyme bovine carbonic anhydrase to neutral -NHCOCH<sub>3</sub> groups makes virtually no difference to the refolding rate or catalytic activity of the protein following denaturation. This is despite the fact that electrostatic repulsion between the NH<sub>3</sub><sup>+</sup> groups is likely to be far smaller in the denatured, extended polypeptide chain.

So why are the charged groups there at all? Might they prevent aggregation? What do they do to the energetics of hydration?

Philip Ball

# Elephants are capable of vocal learning

Two animals coin unexpected sounds as a surprising form of social communication.

There are a few mammalian species that can modify their vocalizations in response to auditory experience<sup>1</sup> — for example, some marine mammals use vocal imitation for reproductive advertisement, as birds sometimes do. Here we describe two examples of vocal imitation by African savannah elephants, *Loxodonta africana*, a terrestrial mammal that lives in a complex fission–fusion society<sup>2</sup>. Our findings favour a role for vocal imitation that has already been proposed for primates, birds, bats and marine mammals: it is a useful form of acoustic communication that helps to maintain individual-specific bonds within changing social groupings<sup>3,4</sup>.

In the first case, we recorded imitations of truck sounds made by Mlaika, a ten-year-old adolescent female African elephant living in a semi-captive group of orphaned elephants in Tsavo, Kenya. Trucks were sometimes audible (Fig. 1a) from Mlaika's night stockade, which lay 3 km from the Nairobi–Mombasa highway. Mlaika emitted truck-like sounds (Fig. 1b; audio file in supplementary information) for several hours after sunset, the



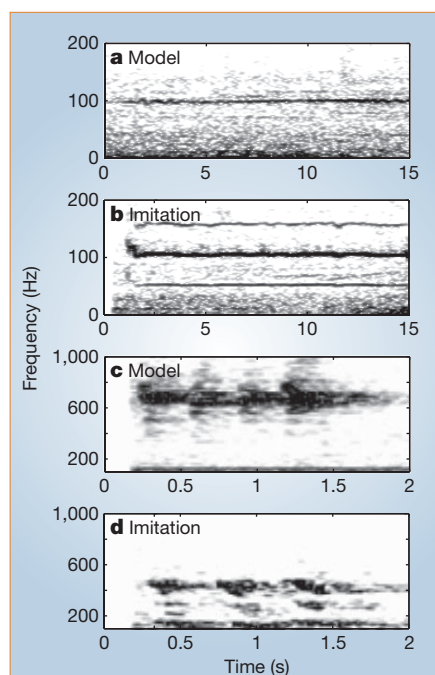
comparisons that used Mlaika's calls or truck sounds, ten samples were randomly selected from the appropriate data set (for methods and further details, see supplementary information).

The individual means for the three acoustic parameters of these calls (Fig. 2b) also show that Mlaika's truck-like calls differ from the normal calls of African elephants<sup>6</sup> and are similar to the recorded truck sounds. Mlaika's truck-like calls are no more similar to the sounds of trucks recorded at the same time than they are to truck sounds recorded at other times ( $t$ -test,  $P > 0.6$ ), suggesting that Mlaika used the general features of truck sounds as a model.

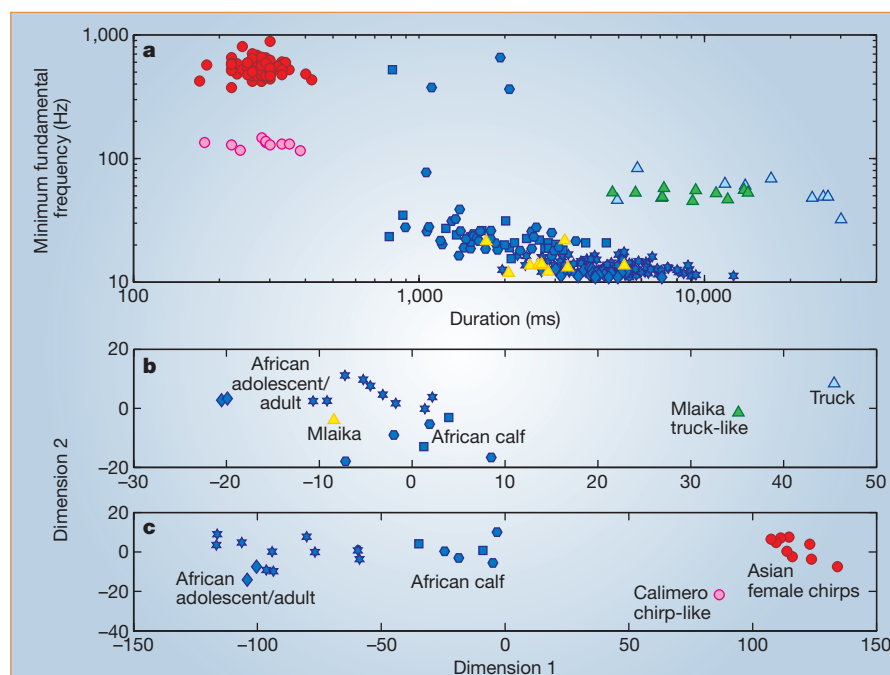
optimal time for the transmission of low-frequency sound in African savannahs<sup>5</sup>.

Comparisons of durations and of minimum and maximum fundamental frequencies show that Mlaika's truck-like calls differ from normal African adult, adolescent and calf calls (Fig. 2a; Mann–Whitney  $U$ -test,  $P < 0.002$ ). Mlaika's truck-like calls do not differ significantly from the truck sounds (Fig. 2a; Mann–Whitney  $U$ -test,  $P > 0.1$ ). For

A second case of imitation by an African elephant involves the chirping sounds<sup>7</sup> typically produced by Asian elephants, *Elephas maximus*, though not by African elephants. Calimero is a 23-year-old male African elephant who spent 18 years with two female Asian elephants at the Basel Zoo in Switzerland. The spectrogram in Fig. 1c shows a typical series of chirps from one of the Asian elephants and Fig. 1d shows the similar



**Figure 1** Spectrograms showing examples of models and sound imitations by two African elephants, Mlaika and Calimero. **a**, Sound of a distant truck recorded from within Mlaika's night stockade; **b**, Mlaika's truck-like call. **c**, Chirps emitted by Asian female elephants living in captivity with Calimero, and **d**, the chirp-like calls emitted by Calimero. (For details and audio files, see supplementary information.)



**Figure 2** Imitation of sounds by the African elephants Mlaika and Calimero. Mlaika's imitations (green triangles) are similar to truck sounds (light-blue triangles) and differ from her normal calls (yellow triangles), which are similar to the sounds made by other African elephants (dark blue symbols: stars, adult female; diamonds, adult male; squares, female calf; hexagons, male calf). Calimero makes chirp-like sounds (pink circles) similar to the chirps of the Asian elephants (red circles) who lived with him. **a**, Scatterplot of frequency versus duration for ten calls from each source. **b**, Multidimensional scaling plot of means for each source in **a**, apart from the chirp sounds. **c**, Multidimensional scaling plot of individual means from chirps of nine female Asian elephants and Calimero's chirp-like calls. Dimension numbering represents different combinations of acoustic features; for methods and further details, see supplementary information.



chirp-like calls made by Calimero, who seldom makes any other sounds.

Using the same statistical analysis as that applied to Mlaika's calls, we find that Calimero's chirp-like calls do not significantly differ in duration from the chirping calls of Asian elephants ( $P > 0.46$ ) and differ in all parameters from mean adult, adolescent and calf African elephant calls ( $P < 0.002$ ; Fig. 2a). Multidimensional scaling confirms that Calimero's chirp-like calls differ from the normal calls of African elephants and are similar to the chirps of female Asian elephants (Fig. 2c).

The evolution of vocal imitation in humans, some birds, bats and dolphins may result from the need to use acoustic signals to maintain individual-specific bonds when animals separate and reunite<sup>1,3</sup>. If so, then vocal learning should occur in other species in which long-lived social bonds are based upon individual-specific relationships and involve fluid group membership, and where vocal communication is used to maintain contact and individual or group recognition. Our finding that an African savannah elephant matched the calls of Asian elephants with whom he lived follows a pattern commonly seen in species that are capable of vocal learning, in which calls converge as the animals form social bonds<sup>3</sup>.

Vocal learning enables a flexible and open communication system in which animals may learn to imitate signals that are not typical of the species, as demonstrated by the elephant that imitated trucks. To our knowledge, this discovery in elephants is the first example of vocal imitation in a non-primate terrestrial mammal. It strengthens the idea that there is a primary selection pressure for vocal learning that involves the communicative demands of maintaining social relationships in fluid societies.

Joyce H. Poole\*, Peter L. Tyack†, Angela S. Stoeger-Horwath‡, Stephanie Watwood†

\*ElephantVoices, Amboseli Trust for Elephants, PO Box 15135, Langata 00509, Nairobi, Kenya

†Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA e-mail: ptyack@whoi.edu

‡Institute of Zoology, University of Vienna, 1090 Vienna, Austria

1 Janik, V. M. & Slater, P. J. B. *Adv. Stud. Behav.* **26**, 59–99 (1997).

2 Moss, C. J. & Poole, J. H. in *Primate Social Relationships: An Integrated Approach* (ed. Hinde, R. A.) 315–325 (Blackwell Scientific, Oxford, 1983).

3 Tyack, P. in *Animal Social Complexity: Intelligence, Culture, and Individualized Societies* (eds de Waal, F. & Tyack, P. L.) 342–361 (Harvard Univ. Press, Cambridge, 2003).

4 McComb, K., Moss, C., Sayialel, S. & Baker, L. *Anim. Behav.* **59**, 1103–1109 (2000).

5 Garstang, M., Larom, D., Raspet, R. & Lindeque, M. J. *Exp. Biol.* **198**, 939–951 (1995).

6 Poole, J. H., Payne, K. B., Langbauer, W. Jr & Moss, C. J. *Behav. Ecol. Sociobiol.* **22**, 385–392 (1988).

7 McKay, G. M. *Smithson. Contrib. Zool.* **125**, 1–113 (1973).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

## Meteorology

# Dusty ice clouds over Alaska

Particles lofted into the atmosphere by desert dust storms can disperse widely and affect climate directly through aerosol scattering and absorption. They can also affect it indirectly by changing the scattering properties of clouds and, because desert dusts are particularly active ice-forming agents, by affecting the formation and thermodynamic phase of clouds. Here I show that dust storms that occurred in Asia early in 2004 created unusual ice clouds over Alaska at temperatures far warmer than those expected for normal cirrus-cloud formation.

Although the direct light-scattering effects of dust storms are readily observed from satellites<sup>1</sup>, their indirect effects on clouds are difficult to quantify. Knowledge about the effects of desert dusts on clouds has come from polarization lidars (detection systems that work on radar principles but use laser light), which can identify both non-spherical aerosols and the phase of clouds<sup>2</sup>. The evidence suggests that, even after great transport distances, aerosols from both the Asian and Saharan deserts affect cirrus and supercooled water clouds<sup>3–6</sup>. It has long been known that desert dusts are highly effective as nuclei on which ice crystals form heterogeneously<sup>7</sup>.

From late February to mid May 2004, ruby (0.694  $\mu\text{m}$  wavelength) polarization-lidar observations<sup>8</sup> from the University of Alaska at Fairbanks (64.86°N, 147.84°W) regularly detected elevated dust layers, which were tracked back to a series of Asian dust

storms. In many cases, the dust layers sporadically developed embedded ice-crystal clouds. This phenomenon is illustrated in Fig. 1 by the backscattered laser power and linear depolarization ratio ( $\delta$ , which is the ratio of the powers in the planes of polarization orthogonal and parallel to that transmitted). Figure 1 shows two hours of lidar display over a range of heights. Aerosols are particularly apparent in stratified layers at between about 4.5 and 6.5 km of height because of their enhanced backscattering (an increase in display brightness) and depolarization (values of  $\delta$ , 0.1–0.2) compared with that produced by molecular scattering. Also prominent is the embedded ice cloud centred at 5.5 km at about 20:30 GMT, which displays even stronger backscattering and depolarization. This cloud had a minimum temperature of  $-36^\circ\text{C}$  and a maximum relative humidity (with respect to ice) of 103% (according to the bracketing local radiosonde data).

Similar mixed dust and ice layers were observed with cloud-top temperatures between about  $-15^\circ\text{C}$  and  $-40^\circ\text{C}$ : these are abnormally warm compared with those of cirrus clouds, which form by the homogeneous nucleation of haze particles<sup>8</sup>. The corresponding relative humidities indicate an almost ice-saturated environment, which also contrasts with the high supersaturations needed for cirrus formation<sup>9</sup>. In laboratory experiments, dust particles needed ice supersaturations of only 10–15% to nucleate ice crystals directly from the vapour<sup>10</sup>.

Although there have been previous indications that desert dusts alter clouds<sup>3–6</sup>, to my knowledge this is the first time that they have been shown to serve as deposition ice nuclei in the atmosphere and to create ice clouds at modest supersaturations and temperatures. Desert dust storms are a natural part of our world, but we must be wary in case human activities in arid lands, or climate change itself, are increasing aerosol amounts and so further altering climate by affecting distant clouds.

Kenneth Sassen

Geophysical Institute, University of Alaska Fairbanks, Fairbanks, Alaska 99775, USA

e-mail: ksassen@gi.alaska.edu

1 Kaufman, Y. K., Tanré, D. & Boucher, O. *Nature* **419**, 215–223 (2002).

2 Sassen, K. in *Lidar* (ed. Weitkamp, C.) 19–42 (Springer, Berlin, 2005).

3 Murayama, T. N. et al. *J. Geophys. Res.* **106**, 18345–18360 (2001).

4 Sassen, K. *Geophys. Res. Lett.* **29**, 1465

doi:10.1029/2001GL014051 (2002).

5 Sassen, K., DeMott, P. J., Prospero, J. M. & Poellot, M. R. *Geophys. Res. Lett.* **30**, 1633 doi:10.1029/2003GL017371 (2003).

6 Sakai, T., Nagai, T., Nakazato, M. & Matsumura, T. *Geophys. Res. Lett.* **31**, L06128 doi:10.1029/2003GL019332 (2004).

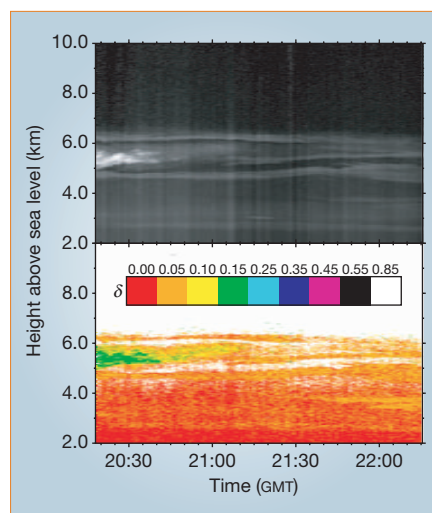
7 Isono, K., Komabayashi, M. & Ono, A. *J. Meteorol. Soc. Jpn* **37**, 211–233 (1959).

8 Sassen, K. & Campbell, J. R. *J. Atmos. Sci.* **58**, 481–496 (2001).

9 Comstock, J. M., Ackerman, T. P. & Turner, D. D. *Geophys. Res. Lett.* **31**, L11106 doi:10.1029/2004GL019705 (2004).

10 Bailey, M. & Hallett, J. Q. *J. R. Meteorol. Soc.* **128**, 1461–1483 (2002).

Competing financial interests: declared none.



**Figure 1** Polarization-lidar height against time display, showing ice-cloud formation in an elevated desert-dust aerosol layer. Displays of backscattered laser power (top; logarithmic grey scale) and linear depolarization ratio (colour scale;  $\delta$  is the ratio of the powers in the planes of polarization orthogonal and parallel to that transmitted) are given for a two-hour period on 29 February 2004, as debris from an Asian dust storm drifted over the interior of Alaska.



# Foreshock sequences and short-term earthquake predictability on East Pacific Rise transform faults

Jeffrey J. McGuire<sup>1</sup>, Margaret S. Boettcher<sup>2</sup> & Thomas H. Jordan<sup>3</sup>

<sup>1</sup>Department of Geology and Geophysics, Woods Hole Oceanographic Institution, and <sup>2</sup>MIT-Woods Hole Oceanographic Institution Joint Program, Woods Hole, Massachusetts 02543-1541, USA

<sup>3</sup>Department of Earth Sciences, University of Southern California, Los Angeles, California 90089-7042, USA

**East Pacific Rise transform faults are characterized by high slip rates (more than ten centimetres a year), predominately aseismic slip and maximum earthquake magnitudes of about 6.5. Using recordings from a hydroacoustic array deployed by the National Oceanic and Atmospheric Administration, we show here that East Pacific Rise transform faults also have a low number of aftershocks and high foreshock rates compared to continental strike-slip faults. The high ratio of foreshocks to aftershocks implies that such transform-fault seismicity cannot be explained by seismic triggering models in which there is no fundamental distinction between foreshocks, mainshocks and aftershocks. The foreshock sequences on East Pacific Rise transform faults can be used to predict (retrospectively) earthquakes of magnitude 5.4 or greater, in narrow spatial and temporal windows and with a high probability gain. The predictability of such transform earthquakes is consistent with a model in which slow slip transients trigger earthquakes, enrich their low-frequency radiation and accommodate much of the aseismic plate motion.**

On average, before large earthquakes occur, local seismicity rates show a significant increase<sup>1</sup>. In continental regions, where dense regional seismic networks provide the best data, most foreshock studies<sup>2–4</sup>, though not all<sup>5</sup>, are consistent with the hypotheses that earthquake nucleation is independent of magnitude and that foreshocks result from a general triggering process in which there is no fundamental distinction between foreshocks, mainshocks and aftershocks. The inability to distinguish foreshocks from the statistical fluctuations in the continental background seismicity severely limits their usefulness in predicting large earthquakes<sup>6</sup>.

It is unclear, however, whether these statements apply to other tectonic environments, or how aseismic processes affect earthquake triggering. Aseismic slip transients with timescales of days to months have recently been observed in the subduction zones of Japan<sup>7–9</sup> and Cascadia<sup>10</sup>, using continuously monitored GPS arrays. The possibility that aseismic slip triggers large earthquakes on subduction megathrusts is especially intriguing given the observation<sup>11</sup> that a slow slip transient occurred 15 minutes before the great 1960 Chilean megathrust earthquake, which had a moment magnitude ( $M_w$ ) of 9.5, the largest ever recorded. Notably, subduction zones are observed to have higher foreshock rates than continental regions<sup>12</sup>.

Another tectonic environment in which aseismic processes are thought to exert a strong influence on fault behaviour is mid-ocean ridge transform faults (RTFs). Studies over the last several decades<sup>13–15</sup> have shown that on average most of the slip on RTFs, up to about 85% (ref. 15), is aseismic. Moreover, the seismic component of slip occurs in earthquakes that are relatively small ( $M_w \leq 7.2$ ) given the length of the faults<sup>14–16</sup>. Many of the larger RTF earthquakes are slow events with anomalous low-frequency radiation<sup>17,18</sup>. Low-frequency spectral analyses<sup>19,20</sup> have indicated that slow RTF earthquakes are compound events comprising an ordinary rupture and a slow transient of comparable moment but much longer duration; in some cases, the slow component precedes, and presumably initiates, the main seismic component. Time-domain records of slow precursors to RTF earthquakes<sup>20,21</sup> and episodes of coupled seismic slip observed on adjacent RTFs<sup>20,22</sup>

support the inference of slow slip transients, but the subject remains controversial<sup>23</sup>.

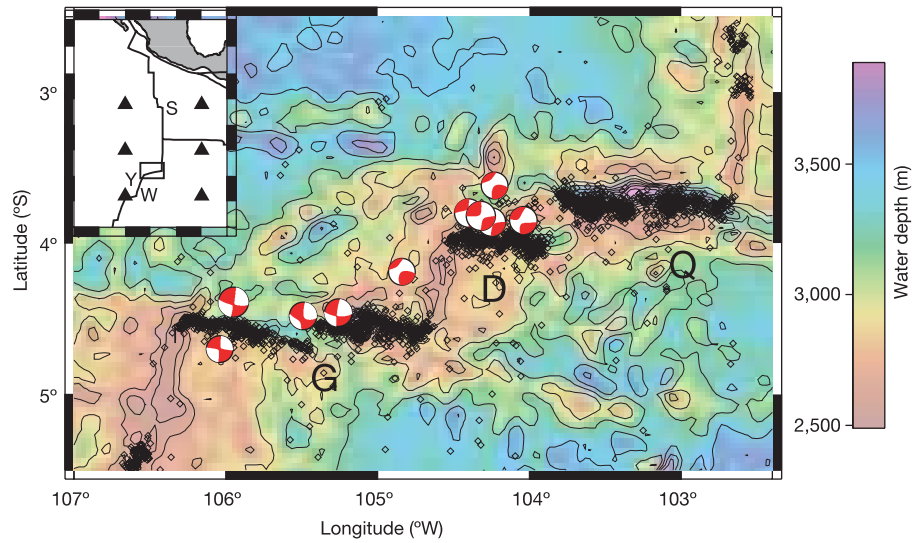
## Hydroacoustic detection of foreshocks

Here we use data from a six-element hydroacoustic array deployed by the National Oceanic and Atmospheric Administration's Pacific Marine Environmental Laboratory (NOAA-PMEL) to examine anomalous foreshock sequences on East Pacific Rise (EPR) transform faults (Fig. 1). The NOAA-PMEL arrays<sup>24–26</sup> routinely locate EPR earthquakes with acoustic source level (ASL) magnitudes (see Methods)  $M_{ASL}$  below 3, reducing the detection threshold by 1.5–2.0 magnitude units below global seismicity catalogues (see Methods). A reconnaissance study indicated that foreshocks in the last hour before large events are significantly more common on EPR transform faults than on strike-slip faults in the North Atlantic, Northeast Pacific or Southern California<sup>27</sup>.

Figure 2 displays stacks of the seismicity in space-time windows centred on nine mainshocks that occurred on the Discovery and Gofar transform faults between May 1996 and December 2001. This set of mainshocks comprised all  $M_w \geq 5.4$  earthquakes on these two faults recorded by the NOAA-PMEL array in the Harvard Centroid Moment Tensor (CMT) catalogue<sup>28</sup> that did not follow within 1 week and 100 km of another mainshock. The longer window (Fig. 2a) shows low background seismicity tens of hours before the mainshocks and the subsequent aftershock decay. The shorter window (Fig. 2b) reveals an accelerating rate of seismicity close to the mainshock epicentres during the hour immediately preceding the mainshock origin times.

## Earthquake triggering model

The anomalous nature of the RTF foreshock activity can be quantified in terms of the Epidemic Type Aftershock Sequence (ETAS) model of triggered seismicity<sup>4,29,30</sup>. ETAS is a marked point process model<sup>31</sup> in which all earthquake magnitudes above a lower cutoff  $m_0$  are independent samples of the Gutenberg–Richter (GR) probability distribution,  $P(m) = 10^{-b(m-m_0)}$ , where  $b$  is the slope of the distribution, and all earthquakes give birth to daughter events at



**Figure 1** Map of the Quebrada (Q), Discovery (D), and Gofar (G) transform faults in the equatorial eastern Pacific, contoured with the bathymetry predicted from the satellite-derived gravity field<sup>42</sup>. Diamond symbols represent the acoustic radiator positions in the NOAA-PMEL seismicity catalogue for 1996–2001<sup>24</sup>, and beachball symbols show the

focal mechanisms and centroid locations for the same period from the Harvard CMT catalogue. The inset map locates the EPR and Cocos ridge crests (black lines), the six NOAA hydrophones (triangles), the Wilkes (W), Yaquina (Y), and Siqueiros (S) transform faults, and the region of the main map (rectangle). The contour interval is 200 m.

an average rate of  $\phi(m, t) = \rho(m)\psi(t)$ , where  $t$  is time since the earthquake. This triggering rate is assumed to increase exponentially with magnitude,  $\rho(m) = k \times 10^{\alpha(m-m_0)}$ , where  $\alpha$  is the triggering exponent, and to decay with time after a mother event according to the modified Omori law,  $\psi(t) = \theta c^\theta / (c + t)^{1+\theta}$  (where  $\theta > 0$ ). The constants  $k$ ,  $\theta$  and  $c$  are parameters that vary among regions. Renormalization for a single mainshock of magnitude  $m$  yields an average seismicity rate proportional to<sup>29</sup>  $N_a^{\text{II}}(m) = \frac{k}{1-n} 10^{\alpha(m-m_0)}$ .  $N_a^{\text{II}}$  is the expected number of aftershocks of any magnitude (type-II aftershocks; that is, not constrained to be smaller than  $m$ ), and the constant  $n = \int_0^1 \rho(\mu) dP(\mu) = kb/(b-\alpha)$  is the branching ratio, which equals both the average number of directly triggered aftershocks per event and the fraction of the earthquake population

that is made up of triggered earthquakes<sup>30</sup>. The aftershock rate decays with an effective Omori exponent<sup>4</sup>  $p = 1 + O(\theta)$ .

In the ETAS model, the seismicity rate before a mainshock at  $t = 0$  increases according to the inverse Omori law; that is,  $\sim |t|^{-p'}$ , where  $p' = 1 + O[\theta]$ , and the expected number of events of all magnitudes conditioned on the mainshock occurrence (type-II foreshocks) is independent<sup>4</sup> of  $m$ . This conditional foreshock number can be approximated as the product of two factors: the probability that the mainshock is a triggered event, and the expected number of events in a cluster averaged over mainshock magnitude. The first is just the branching ratio  $n$ , and the second is the integral  $\int_0^1 N_a^{\text{II}}(\mu) dP(\mu)$ ; therefore<sup>30</sup>,  $N_f^{\text{II}} \approx n^2 / (1 - n)$ . To include only earthquakes smaller than the mainshock (type-I foreshocks), we multiply the integrand by the probability that no event in a cluster exceeds  $m$  and integrate over the appropriate magnitude range. If  $k/(1 - n)$  is small and  $m$  is large (conditions which apply to our data), then the extra probability factor is close to unity, and the results are  $N_f^{\text{I}} \approx N_f^{\text{II}} [1 - 10^{-(b-\alpha)(m-m_0)}]$ .

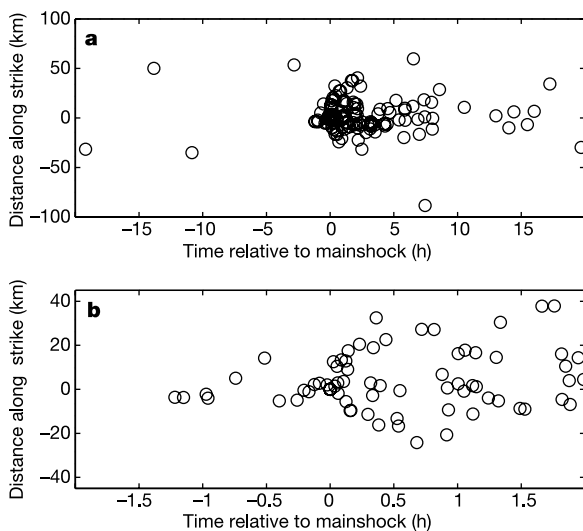
A similar modification to the aftershock number yields the foreshock/aftershock ratio:

$$\frac{N_f}{N_a} \approx n \left( \frac{b}{b-\alpha} \right) \left[ \frac{10^{(b-\alpha)\Delta m_1^f} - 10^{(b-\alpha)\Delta m_2^f}}{10^{b\Delta m_1^a} - 10^{b\Delta m_2^a}} \right] \quad (1)$$

Here we have generalized the formula to count foreshocks in the magnitude range from  $m - \Delta m_1^f$  to  $m - \Delta m_2^f$  and aftershocks from  $m - \Delta m_1^a$  to  $m - \Delta m_2^a$ , where  $0 \leq \Delta m_2^{f,a} < \Delta m_1^{f,a} \leq m - m_0$ . This approximation, which applies to large mainshocks, differs conceptually from the expression recently used by ref. 3 to explain the foreshock/aftershock ratios from global and regional catalogues (see Methods).

### Anomalous foreshock activity

Earthquake populations on RTFs are well described by a tapered GR distribution having a low-magnitude slope  $b \approx 1$  (refs 14, 15) similar to that of continental regions. The hydroacoustic catalogue for the EPR faults is consistent with this self-similar scaling, and its aftershock sequences decay according to Omori's law with<sup>32</sup>  $p \approx 1$ , again similar to continental regions. However, global catalogues demonstrate that the aftershock productivity of large RTF earthquakes is lower than continental faults by approximately a factor of fifteen<sup>15</sup> (Fig. 3). The low aftershock productivity combined with

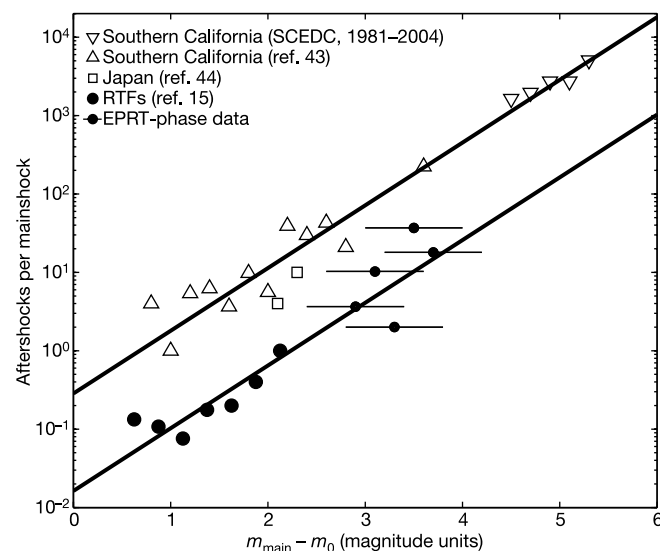


**Figure 2** Space-time distribution of seismicity around the nine mainshocks ( $M_w \geq 5.4$ ) on the Discovery and Gofar transform faults between May 1996 and December 2001, from the declustered Harvard CMT catalogue. **a**, Stack of all events from the NOAA-PMEL hydroacoustic catalogue with  $M_{\text{ASL}} > 2.5$  (for  $\text{ASL} > 207$ ) that were located within  $\pm 100$  km along strike and within  $\pm 20$  h of the mainshocks. Positive distance is west of the mainshock, and positive time is after the mainshock. **b**, Zoomed-in view of the same seismicity, showing foreshock activity within about 1 h and 15 km of the mainshocks.

the poor detection thresholds of global catalogues makes it difficult to constrain the values of  $n$  and  $\alpha$  independently. A maximum-likelihood fit to the teleseismic RTF data yields a best-fit value of the triggering exponent  $\alpha = 0.72$ , and is consistent with the somewhat higher values found for California and Japan ( $\alpha = 0.8\text{--}1.0$ )<sup>33,34</sup> (Fig. 3). Error bounds on the maximum-likelihood estimate are large, but aftershock counts using the hydroacoustic catalogue (points with horizontal bars in Fig. 3) also favour relatively high values of  $\alpha$  and rule out values less than about 0.6 (see Supplementary Information).

The difference between oceanic and continental aftershocks primarily manifests itself in the intercept of the scaling relation,  $k/(1 - n)$ , which is offset by about a factor of fifteen (Fig. 3). The maximum-likelihood fit in Fig. 3 corresponds to a branching ratio,  $n \approx 0.1$ , compared to values approaching unity in continental seismic zones<sup>35</sup>. As discussed in the Supplementary Information, the aftershock rate may be somewhat higher for the EPR faults, but we can say with a high degree of confidence that  $n < 0.3$ . In other words, according to the ETAS model, most RTF earthquakes (70–90%) would be primary events driven by aseismic plate-tectonic loading rather than aftershocks of previous earthquakes. Equation (1) with the maximum-likelihood estimate of  $\alpha$  and  $n$  predicts that the foreshock/aftershock ratio for RTFs should be about an order of magnitude lower than that observed in continents.

Instead, the EPR transform faults that are well recorded by the NOAA-PMEL array give values of  $N_f/N_a$  that are an order of magnitude higher than observed in Southern California (Fig. 4). For both regions, we identified mainshocks as events in the Harvard CMT catalogue with  $M_w \geq 5.4$  that did not follow within 1 week and 100 km of another mainshock (see Methods), and we compiled

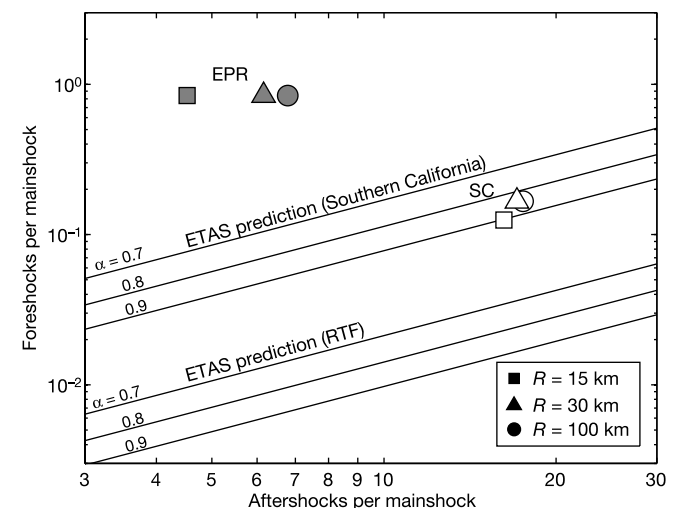


**Figure 3** Aftershocks per mainshock, plotted against the difference between the mainshock magnitude  $m_{\text{main}}$  and the catalogue completeness threshold  $m_0$ . RTF aftershocks (large filled circles) were defined as events with calibrated surface-wave magnitudes above  $m_0 = 5.1$  that occurred within 14 days and 100 km of a  $M_w \geq 5.6$  mainshock during the catalogue interval 1976–2001 (ref. 15). Southern California aftershocks from the SCEDC catalogue (open triangles) were defined as events above a local-magnitude ( $M_L$ ) threshold of  $m_0 = 2.0$  that occurred within 14 days and 100 km of a  $M_L \geq 6.5$  mainshock during the interval 1981–2004. Aftershock counts from the EPR  $T$ -phase catalogue (small filled circles) are shown with error bars to account for uncertainties in  $m_0$  ( $2.0 \leq m_0 \leq 3.0$ ). The  $T$ -phase catalogue aftershocks were counted within 14 days and 30 km of the mainshocks. Previously published continental data sets (open triangles and squares) were compiled by Kisslinger and Jones<sup>43</sup> and Yamanaka and Shimazaki<sup>44</sup> using  $M_L$  thresholds of  $m_0 = 4.0$  and  $4.5$ , respectively. Both RTF and continental aftershocks are consistent with a triggering exponent of  $\alpha = 0.8$  (solid lines), but RTFs produce fewer aftershocks by a factor of fifteen.

foreshock and aftershock statistics from the NOAA-PMEL and the Southern California Earthquake Data Center (SCEDC) catalogues. We counted all events with local magnitudes ( $M_{\text{SL}}$  or  $M_L$ ) up to 2.8 units smaller than the mainshock  $M_w$  in the 1-h interval before and the 5-h interval after the mainshock. Figure 4 compares the observed  $N_f/N_a$  for spatial windows of various radii with the predictions of equation (1), corrected for the finite sampling intervals (see Methods). The SCEDC statistics satisfy an ETAS model with  $\alpha = 0.8\text{--}0.9$ , consistent with previous catalogue studies<sup>3,29</sup>. However, foreshock rates from the NOAA-PMEL statistics are about two orders of magnitude greater than the ETAS predictions using the maximum-likelihood fit in Fig. 3. As shown in Fig. 4 and Supplementary Fig. S3, these results are robust with respect to the choice of windows and declustering procedures.

Therefore, we can reject the ETAS hypothesis that the clustering of foreshocks, mainshocks and aftershocks on RTFs can be described by the same seismic triggering mechanism. We infer that large earthquakes on EPR faults are preceded by an extended preparation process, possibly driven by subseismic transients (silent or quiet earthquakes), that can often be observed through foreshocks. This alternative hypothesis is consistent with the tightly localized distribution of the foreshocks about the mainshock in both space and time (Fig. 2 and Supplementary Fig. S4), which does not conform to the inverse-diffusive behaviour expected from the ETAS model<sup>4</sup>.

The correspondence of slow slip with foreshocks was suggested as early as 1976 by Kanamori and Stewart<sup>18</sup>, who noted a foreshock with a body-wave magnitude  $m_b \approx 5$  about 500 s before the  $M_w = 7$  slow earthquake on the Gibbs transform fault in the North Atlantic. More recently, McGuire *et al.* associated  $m_b = 4.5\text{--}5.0$  foreshocks before the 1994  $M_w = 7.0$  Romanche<sup>20</sup> and 1997  $M_w = 6.8$  Prince Edward Island<sup>21</sup> earthquakes with slow precursors observed at low frequencies. Forsyth *et al.*<sup>22</sup> suggested



**Figure 4** Foreshock and aftershock rates observed for EPR transform faults (solid symbols) and Southern California (open symbols) in regions of radius  $R$  about the mainshock. The data sets included 19 mainshocks ( $M_w \geq 5.4$ ) on five transform faults (Discovery, Gofar, Wilkes, Yaquina and Siqueiros) from the declustered Harvard CMT catalogue for 1996–2001, and 24 mainshocks ( $M_L \geq 5.4$ ) in Southern California from the declustered SCEDC catalogue for 1981–2003. Events with magnitudes up to 2.8 units below the mainshock magnitude were counted from the NOAA-PMEL and SCEDC catalogues in the 1-h window preceding and 5-h window following the mainshocks. These rates are compared with the  $N_f/N_a$  ratios from the ETAS model (equation (1)) for  $\alpha = 0.7\text{--}0.9$  (solid lines), assuming  $\Delta m_2^{\text{f,a}} = 0$ ,  $\Delta m_1^{\text{f,a}} = 2.8$  and  $b = 1$ , with estimated branching ratios of  $n = 0.8$  (Southern California) and  $n = 0.1$  (RTF). The  $\alpha = 0.8$  line for RTFs is close to the maximum-likelihood estimate from Fig. 3. Uncertainties in  $\alpha$  and  $n$  allow shifts in the ETAS prediction upwards from the maximum-likelihood value by half an order of magnitude at most.



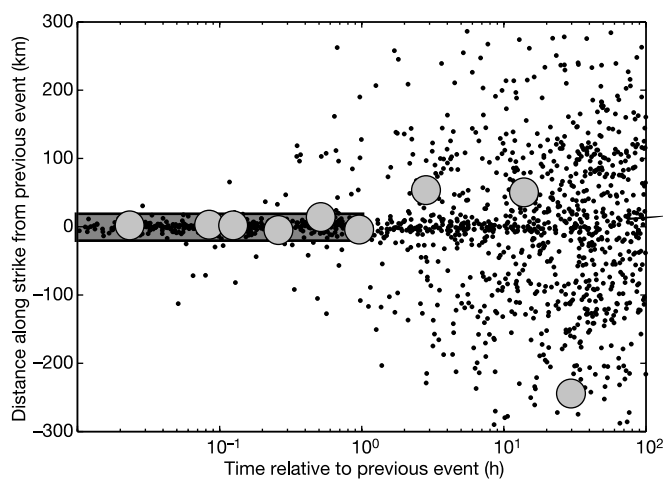
that a subseismic slip process was responsible for a swarm of contemporaneous seismicity on the Anakena and Raraku transform faults of the southern EPR recorded by an ocean-bottom seismometer array in 1995. These and other examples<sup>36</sup> combined with the global aftershock depletion (Fig. 3) and the evidence for slow precursors to large earthquakes on RTFs worldwide<sup>19</sup>, indicate that the aseismic, foreshock-generating process on EPR faults may be prevalent throughout the mid-ocean-ridge system, including the slower-slipping, colder RTFs in the Atlantic and Indian oceans.

### Short-term predictability of large earthquakes

The high rate of proximate foreshocks suggests a naive scheme for short-term earthquake prediction—we simply assume that every event is a foreshock of an impending large earthquake. We can formalize this scheme into a well-posed prediction algorithm<sup>37</sup>: whenever we observe any RTF event above some ASL magnitude threshold  $m_0$  within a specified RTF region, we issue an alert that an earthquake of moment magnitude greater than or equal to  $m_p$  will occur sometime during time window of length  $t_p$  immediately following the event and somewhere in a spatial window of radius  $r_p$  about the event's epicentre. Figure 5 illustrates this prediction algorithm for the parameter set [ $m_0 = 2.5$  ( $M_{ASL}$ ),  $m_p = 5.4$  ( $M_W$ ),  $t_p = 1$  h,  $r_p = 15$  km] by applying it retrospectively to the two most active EPR transform faults, Discovery and Gofar. Of the nine candidate earthquakes that occurred during the catalogue interval May 1996–November 2001, six were located within the space-time prediction windows (Fig. 5) and thus constitute successful predictions. There were three failures-to-predict and about 1,400 false alarms.

Although the false-alarm rate is quite high, all alarms taken together occupy only about 0.15% of the total space-time volume of about 250 km  $\times$  5.5 yr (see Supplementary Information). We can relate  $P(M|F)$ , the probability of a mainshock  $M$  in the prediction window given the occurrence of a foreshock  $F$ , to  $P(M)$ , the probability of  $M$  in a random window of the same size, using the Bayes identity:

$$P(M|F) = P(M) \left[ \frac{P(F|M)}{P(F)} \right] \quad (2)$$



**Figure 5** Retrospective application of the naive prediction algorithm described in the text to the NOAA-PMEL catalogue (May 1996–November 2001) for the Discovery and Gofar faults. Plot shows along-strike distance (positive to the west) and time of each earthquake relative to its previous event for all catalogued events with  $M_{ASL} \geq 2.5$  ( $ASL \geq 207$ ). Events were considered to be distinct if they were separated by more than 1 min from the previous event, to exclude redundancies. Six of the nine mainshocks identified from the declustered Harvard CMT catalogue (large circles) fall within the 1-h,  $\pm 15$ -km alert windows (shaded area) used in the prediction algorithm.

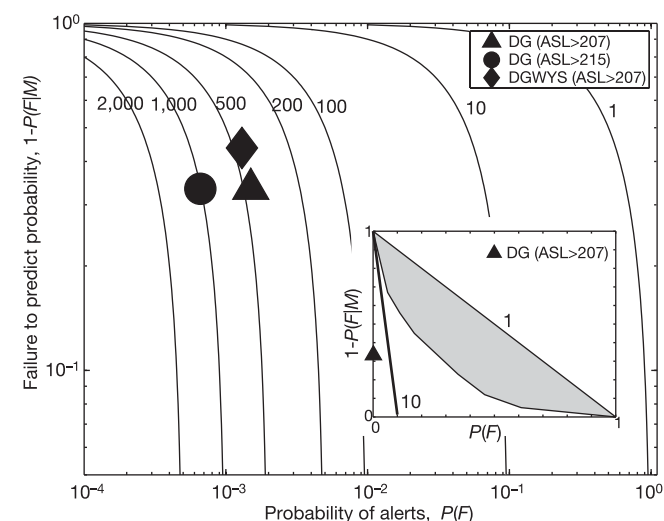
$P(F|M)$  is the fraction of mainshocks preceded by foreshocks, and  $P(F)$  equals the fraction of the space-time volume occupied by alerts. The ratio of these probabilities (the term in square brackets) is the probability gain factor  $g$  of the prediction algorithm<sup>37</sup>. Our retrospective analysis of the Discovery and Gofar faults gives  $g = 450$ . This performance can be compared to prediction experiments in California, where even highly optimized algorithms with many adjustable parameters are thought to achieve probability gains of 10–20 or less<sup>38</sup>.

The performance relative to random chance can be evaluated using Molchan's<sup>39</sup> error diagram, which plots the failure-to-predict probability,  $1 - P(F|M)$ , against  $P(F)$  (Fig. 6). Large sets of random alerts should fall close to the line  $1 - P(F)$ , corresponding to no probability gain ( $g = 1$ ). Figure 6 shows that the Discovery–Gofar point lies well below the 99% confidence range for random chance; the probability of reproducing this performance with random alerts filling 0.15% of the space-time volume is less than one in ten million. Similar results were obtained by applying the algorithm retrospectively to the five active EPR transforms in the study area; nine of sixteen mainshocks were successfully predicted by alerts occupying 0.13% of the space-time volume, which gives  $g = 340$  (Fig. 6, see Supplementary Discussion).

Our naive algorithm is far from optimal. For instance, raising the threshold magnitude  $m_0$  from 2.5 to 3.4 (that is, increasing the ASL from 207 to 215) reduces the number of alarms in the Gofar–Discovery catalogue to 407 without changing  $P(F|M)$ , increasing  $g$  to about 1,000 (Fig. 6). More parameters could be added to improve the performance further.

### Discussion

Mid-ocean ridges are far removed from urban centres on continents, so the direct societal value of short-term earthquake prediction on RTFs (assuming it could be operationally implemented using real-time, near-source monitoring) would be small. Nevertheless, the existence of short-term predictability in this tectonic environment—the main conclusion of this paper—is of considerable scientific interest, because it supports a physical linkage between foreshocks and mainshocks through stress changes driven by aseismic slip transients or some other fault preparation process



**Figure 6** Molchan's<sup>39</sup> error diagram of the failure-to-predict probability  $1 - P(F|M)$  against the probability of alerts  $P(F)$  on a logarithmic scale, contoured with probability gain  $g$  (solid curves). Performance of the naive prediction algorithm is given for the Discovery and Gofar (DG) faults with  $m_0 = 2.5$  (triangle) and  $m_0 = 3.4$  (circle), and for the five active EPR faults (DGWYS) with  $m_0 = 2.5$  (diamond). Inset diagram is the same plot on a linear scale, comparing the 99% confidence region for a random prediction of nine mainshocks (shaded area) with the first DG test (triangle).

such as hydrothermal flow or nearby magmatic activity.

To monitor RTFs geodetically will require a new generation of ocean-bottom instrumentation, a considerable technological challenge. But collecting such data may well be the best way to test the hypothesis that three anomalous aspects of RTF seismicity—large slip deficits, high foreshock activity and slow earthquakes—can be explained by aseismic fault-slip transients. Given the importance of understanding the fundamental mechanics of earthquake predictability, overcoming the technological hurdles should be worth the effort. □

## Methods

We use the NOAA-PMEL hydroacoustic earthquake catalogue for the equatorial Pacific which began in 1996 (ref. 24). Their array (Fig. 1) records acoustic energy radiated into the water column by earthquakes and other sources ( $T$ -phases).  $T$ -phases propagate very efficiently in the low-velocity sound fixing and ranging (SOFAR) channel, and the array arrivals can be used to locate precisely where the energy entered the SOFAR channel. The standard errors in this source location are estimated to be  $\pm 2$  km,  $\pm 10$  s for the southern/northernmost faults (the Wilkes and Siqueiros faults) and are slightly smaller for the faults located in the centre of the array<sup>24</sup> (the Discovery and Gofar faults). A propagation model is used to convert the magnitude of the pressure signal at the hydrophones into an ASL (measured in decibels) at the source location. We used a conversion between ASL and magnitude,  $M_{\text{ASL}} = 0.107\text{ASL} - 19.6$ , obtained from the regression of the observed ASLs compiled in ref. 32 against the body-wave magnitudes of the International Seismological Centre (ISC) catalogue (<http://www.isc.ac.uk/>). Frequency-magnitude statistics indicate that the seismicity catalogues are approximately complete down to  $\text{ASL} \approx 207$ –212, or  $M_{\text{ASL}} \approx 2.5$ –3.0.

To prevent biases from ongoing aftershock sequences, we eliminated CMT catalogue earthquakes that occurred within 1 week and 100 km of a previous CMT earthquake from our analysis. Our declustering criterion eliminated only one earthquake with  $M_w \geq 5.5$  from the Discovery–Gofar set and only two from the mainshock set for other three active RTFs within the NOAA-PMEL array. Moderate increases in the space and time windows did not disqualify additional events. Moreover, each of the three disqualified events had a foreshock sequence distinct from the aftershock sequence of the preceding mainshock, consistent with the statistics in Fig. 4 (see Supplementary Fig. 2).

The time dependence of foreshock and aftershock rates in ETAS is controlled by Omori's law. Helmstetter *et al.*<sup>4</sup> have demonstrated that the time-decay exponent  $p$  in the renormalized Omori law is not strictly a constant. When  $\rho(m)$  has finite variance ( $\alpha/b \leq 1/2$ ),  $p$  varies from  $1 - \theta$  for  $t \ll t^* = c(n/1 - n)^{1/\theta}$  to  $1 + \theta$  for  $t \gg t^*$ . The theory breaks down for  $\alpha/b > 1/2$ , owing to the strong coupling between earthquake energy and seismicity rate, but the numerical simulations of ref. 4 indicate that the short-time value of  $p$  increases approximately linearly from  $1 - \theta$  at  $\alpha/b = 1/2$  to  $1 + \theta$  as  $\alpha/b \rightarrow 1$ . For  $\alpha/b = 0.8$ , the Omori exponents can be approximated by  $p' \approx p \approx 1 + \theta$ ; hence, the requisite integrals are  $\int_0^{\Delta t_{f,a}} \psi(\tau) d\tau \approx 1 - (c/\Delta t_{f,a})^\theta$ , where the approximation assumes  $\Delta t_{f,a} \gg c$ . Equation (1) can thus be corrected for the finite foreshock and aftershock sampling intervals,  $\Delta t_f = 1$  h and  $\Delta t_a = 5$  h, by multiplying its right-hand side by the ratio of these integrals. This ratio varies from 0.72 to 0.94 over the plausible range of parameters  $c = 1$  s – 1 min,  $\theta = 0$ –0.2, so the correction is minor.

Equation (1) is a more general form of models used in previous studies<sup>3,12,40,41</sup>. Feltzer *et al.*<sup>3</sup> consider the case  $\alpha = b$ . Taking this limit in equation (1), we obtain  $N_f/N_a \approx n(\ln 10) \left[ \frac{\Delta m_f^1 - \Delta m_f^2}{10^{\Delta m_f^1} - 10^{\Delta m_f^2}} \right]$ . For the magnitude ranges of ref. 3 ( $\Delta m_f^1 = 1.0$ ,  $\Delta m_f^2 = 0$ ;  $\Delta m_a^1 = 1.0$ ,  $\Delta m_a^2 = 0.4$ ), the ratio in brackets reproduces their foreshock/aftershock ratio of 0.134. The factor  $n(\ln 10)$  missing in their equation (6) probably lies between 1 and 2 for continental and subduction-zone seismicity<sup>30</sup>, so their formula provides an adequate approximation in most regions. For RTFs, however, the scaling of  $N_f/N_a$  with the branching ratio  $n$  is numerically important (see Fig. 4).

Received 3 September 2004; accepted 19 December 2005; doi:10.1038/nature03377.

- Jones, L. M. & Molnar, P. Some characteristics of foreshocks and their possible relationship to earthquake prediction and premonitory slip on faults. *J. Geophys. Res.* **84**, 3596–3608 (1979).
- Abercrombie, R. E. & Mori, J. Occurrence patterns of foreshocks to large earthquakes in the Western United States. *Nature* **381**, 303–307 (1996).
- Felzer, K., Abercrombie, R. E. & Ekstrom, G. A common origin for aftershocks, foreshocks, and multiplets. *Bull. Seismol. Soc. Am.* **94**(1), 88–98 (2004).
- Helmstetter, A., Sornette, D. & Grasso, J.-R. Mainshocks are aftershocks of conditional foreshocks: How do foreshock statistical properties emerge from aftershock laws. *J. Geophys. Res.* **108**, doi:10.1029/2002JB001991 (2003).
- Dodge, D. A., Ellsworth, W. E. & Beroza, G. C. Detailed observations of California foreshock sequences: Implications for the earthquake initiation process. *J. Geophys. Res.* **101**, 22371–22392 (1996).
- Geller, R. J., Jackson, D. D., Kagan, Y. Y. & Mulargia, F. Earthquakes cannot be predicted. *Science* **275**, 1616–1617 (1997).
- Hirose, H., Hirahara, K., Kimata, F., Fujii, N. & Miyazaki, S. A slow thrust slip event following the two 1996 Hyuganada earthquakes beneath the Bungo Channel, Southwest Japan. *Geophys. Res. Lett.* **26**, 3237–3240 (1999).
- Ozawa, S. *et al.* Detection and monitoring of ongoing aseismic slip in the Tokai region, central Japan. *Science* **298**, 1009–1012 (2002).
- Ozawa, S. *et al.* Characteristic silent earthquakes in the eastern part of the Boso peninsula, Central Japan. *Geophys. Res. Lett.* **30**, doi:10.1029/2002GL016665 (2003).
- Dragert, H., Wang, K. & James, T. S. A silent slip event on the deeper Cascadia subduction interface. *Science* **292**, 1525–1528 (2001).

- Kanamori, H. & Cipar, J. Focal process of the great Chilean earthquake. *Phys. Earth Planet. Inter.* **9**, 128–136 (1974).
- Reasenber, P. A. Foreshock occurrence before large earthquakes. *J. Geophys. Res.* **104**, 4755–4768 (1999).
- Brune, J. N. Seismic moment, seismicity, and rate of slip along major fault zones. *J. Geophys. Res.* **73**, 777–784 (1968).
- Bird, P., Kagan, Y. & Jackson, D. in *Plate Boundary Zones* (ed. Freymueller, S. S. a. J.) 203–218 (AGU, Washington DC, 2002).
- Boettcher, M. S. & Jordan, T. H. Earthquake scaling relations for mid-ocean ridge transform faults. *J. Geophys. Res.* **109**, doi:10.1029/2004JB003110 (2004).
- Bird, P. & Kagan, Y. Plate-tectonic analysis of shallow seismicity: Apparent boundary width, beta, corner magnitude, coupled lithosphere thickness, and coupling in 7 tectonic settings. *Bull. Seismol. Soc. Am.* **94**(6), 2380–2399 (2004).
- Perez-Campos, X., McGuire, J. J. & Beroza, G. C. Resolution of the slow earthquake/apparent stress paradox for oceanic transform fault earthquakes. *J. Geophys. Res.* **108**, doi:10.1029/2002JB002312 (2003).
- Kanamori, H. & Stewart, G. S. Mode of strain release along the Gibbs fracture zone, Mid-Atlantic Ridge. *Phys. Earth Planet. Inter.* **11**, 312–332 (1976).
- Ihmlé, P. F. & Jordan, T. H. Teleseismic search for slow precursors to large earthquakes. *Science* **266**, 1547–1551 (1994).
- McGuire, J. J., Ihmlé, P. F. & Jordan, T. H. Time-domain observations of a slow precursor to the 1994 Romanche transform earthquake. *Science* **274**, 82–85 (1996).
- McGuire, J. J. & Jordan, T. H. Further evidence for the compound nature of slow earthquakes: The Prince Edward Island earthquake of April 28, 1997. *J. Geophys. Res.* **105**, 7819–7828 (2000).
- Forsyth, D. W., Yang, Y., Mangriotes, M.-D. & Shen, Y. Coupled seismic slip on adjacent oceanic transform faults. *J. Geophys. Res.* **30**, doi:10.1029/2002GL016454 (2003).
- Abercrombie, R. E. & Ekstrom, G. Earthquake slip on oceanic transform faults. *Nature* **410**, 74–77 (2001).
- Fox, C. G., Matsumoto, H. & Lau, T. K. Monitoring Pacific Ocean seismicity from an autonomous hydrophone array. *J. Geophys. Res.* **106**, 4183–4206 (2001).
- Fox, C. G. *et al.* Acoustic detection of seafloor spreading. *J. Geophys. Res.* **102**, 131–134 (1995).
- Smith, D. *et al.* Hydroacoustic monitoring of seismicity at the slow-spreading Mid-Atlantic Ridge. *Geophys. Res. Lett.* **29**, 10.1029/2001GL013912 (2002).
- McGuire, J. J. Immediate foreshock sequences of oceanic transform earthquakes on the East Pacific Rise. *Bull. Seismol. Soc. Am.* **93**, 948–952 (2003).
- Ekstrom, G., Dziewonski, A. M., Maternovskaya, N. N. & Nettles, M. Global seismicity of 2001: centroid-moment tensor solutions for 961 earthquakes. *Phys. Earth Planet. Inter.* **136**, 165–185 (2003).
- Helmstetter, A. & Sornette, D. Bath's law derived from the Gutenberg-Richter law and from aftershock properties. *Geophys. Res. Lett.* **30**, doi:10.1029/2003GL018186 (2003).
- Helmstetter, A. & Sornette, D. Importance of direct and indirect triggered seismicity in the ETAS model of seismicity. *Geophys. Res. Lett.* **30**, 4, doi:10.1029/2003GL017670 (2003).
- Ogata, Y. Statistical models for earthquake occurrence and residual analysis for point processes. *J. Am. Stat. Assoc.* **83**, 9–27 (1988).
- Bohnenstiehl, D. R., Tolstoy, M., Dziak, R. P., Fox, C. G. & Smith, D. Aftershocks in the mid-ocean ridge environment: An analysis using hydroacoustic data. *Tectonophysics* **354**, 49–70 (2002).
- Helmstetter, A. Is earthquake triggering driven by small earthquakes? *Phys. Rev. Lett.* **91**, 058501 (2003).
- Helmstetter, A., Kagan, Y. & Jackson, D. D. Importance of small earthquakes for stress transfers and earthquake triggering. *J. Geophys. Res.* (in press).
- Helmstetter, A. & Sornette, D. Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. *J. Geophys. Res.* **107**, doi:10.1029/2001JB001580 (2002).
- Hanson, J. A. & Given, H. K. Accurate azimuth estimates from a large aperture hydrophone array using T-phase waveforms. *Geophys. Res. Lett.* **25**, 365–368 (1998).
- Aki, K. in *Earthquake Prediction* (eds Simpson, D. W. & Richards, P. G.) 556–574 (AGU, Washington DC, 1981).
- Keilis-Borok, V., Shebalin, P., Gabriello, A. & Turcotte, D. Reverse tracing of short-term earthquake precursors. *Phys. Earth Planet. Inter.* **145**, 75–85 (2004).
- Molchan, G. M. Earthquake prediction as a decision-making problem. *Pure Appl. Geophys.* **149**, 233–247 (1997).
- Reasenber, P. A. & Jones, L. M. Earthquake hazards after a mainshock in California. *Science* **243**, 1173–1176 (1989).
- Reasenber, P. A. & Jones, L. M. Earthquake aftershocks: Update. *Science* **265**, 1251–1252 (1994).
- Sandwell, D. T. & Smith, W. H. F. Marine gravity anomaly from Geosat and ERS 1 satellite altimetry. *J. Geophys. Res.* **102**, 10039–10054 (1997).
- Kisslinger, C. & Jones, L. M. Properties of aftershock sequences in Southern California. *J. Geophys. Res.* **96**, 11947–11958 (1991).
- Yamanaka, Y. & Shimazaki, K. Scaling relationship between the number of aftershocks and the size of the main shock. *J. Phys. Earth* **38**, 305–324 (1990).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank R. Dziak for answering questions about details of the hydroacoustic earthquake catalogues, D. Bohnenstiehl for suggestions on clarifying the manuscript, A. Helmstetter for her help in understanding ETAS, and V. Keilis-Borok, I. Zaliapin, and L. Jones for discussions of earthquake prediction algorithms. J.J.McG. was supported by the Frank and Lisina Hoch Fund. M.S.B. was supported by the Deep Ocean Exploration Institute at WHOI. This work was supported by the NSF, SCEE and USGS.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to J.J.McG. (jmcguire@whoi.edu).

# Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*

B. Sönnichsen<sup>1</sup>, L. B. Koski<sup>1\*</sup>, A. Walsh<sup>1</sup>, P. Marschall<sup>1\*</sup>, B. Neumann<sup>1\*</sup>, M. Brehm<sup>1</sup>, A.-M. Alleaume<sup>1\*</sup>, J. Artelt<sup>1\*</sup>, P. Bettencourt<sup>1\*</sup>, E. Cassin<sup>2\*</sup>, M. Hewitson<sup>1</sup>, C. Holz<sup>1</sup>, M. Khan<sup>1</sup>, S. Lazik<sup>1</sup>, C. Martin<sup>1</sup>, B. Nitzsche<sup>1\*</sup>, M. Ruer<sup>2</sup>, J. Stamford<sup>2</sup>, M. Winzi<sup>1</sup>, R. Heinkel<sup>1\*</sup>, M. Röder<sup>1\*</sup>, J. Finell<sup>1\*</sup>, H. Häntschi<sup>1</sup>, S. J. M. Jones<sup>3</sup>, M. Jones<sup>4\*</sup>, F. Piano<sup>5</sup>, K. C. Gunsalus<sup>5</sup>, K. Oegema<sup>2\*</sup>, P. Gönczy<sup>2\*</sup>, A. Coulson<sup>4\*</sup>, A. A. Hyman<sup>2</sup> & C. J. Echeverri<sup>1</sup>

<sup>1</sup>Cenix BioScience GmbH, Tatzberg 47-51, D-01307 Dresden, Germany

<sup>2</sup>Max-Planck-Institute for Cell Biology and Genetics (MPI-CBG), Pfotenhauerstrasse 108, D-01307 Dresden, Germany

<sup>3</sup>Genome Sciences Centre, British Columbia Cancer Research Centre, 570 West 7th Avenue, Vancouver, British Columbia V5Z 4E6, Canada

<sup>4</sup>The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

<sup>5</sup>Department of Biology, Centre for Comparative Functional Genomics, New York University, 1009 Silver Centre, 100 Washington Square East, New York, New York 10003, USA

\* Present addresses: Centre Robert Cedergren, Centre de Recherche en Bioinformatique et en Sciences Génomiques de l'Université de Montréal, 2900 Edouard-Montpetit, Montreal, Quebec H3T 1J4, Canada (L.B.K.); Max Planck Institute of Infection Biology, Schumannstrasse 21/22, 10117 Berlin, Germany (P.M.); European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, D-69117 Heidelberg, Germany (B. Neumann); Institut Cochin, Département de Biologie Cellulaire, 22 rue Méchain, 75014 Paris, France (A.-M.A.); Institut für Klinische Genetik, Medizinische Fakultät Carl Gustav Carus, TU Dresden, Fetscherstrasse 74, 01307 Dresden, Germany (J.A.); Instituto Gulbenkian de Ciência, Rua da Quinta Grande n 6, 2780-901 Oeiras, Portugal (P.B.); Cavaleri Ottolenghi Scientific Institut (COSI) of Neurobiology, Azienda Ospedale San Luigi Gonzaga, Regione Gonzole 10-1004 Orbassano, Italy (E.C.); Mühlstrasse 36, 04317 Leipzig, Germany (B. Nitzsche); Brombeerweg 2, 76275 Ettlingen, Germany (R.H.); Ingeniweb, 2 cours du 14 Juillet, 78300 Poissy, France (M.R.); Rätälänkatu 20 A 11, 20810 Turku, Finland (J.F.); Institute of Molecular Biology and Biochemistry, Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada (M.J.); University of California, San Diego, CMM-East 3080, 9500 Gilman Drive, La Jolla, California 92093, USA (K.O.); Swiss Institute for Experimental Cancer Research (ISREC), CH-1066 Lausanne, Switzerland (P.G.); MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK (A.C.)

**A key challenge of functional genomics today is to generate well-annotated data sets that can be interpreted across different platforms and technologies. Large-scale functional genomics data often fail to connect to standard experimental approaches of gene characterization in individual laboratories. Furthermore, a lack of universal annotation standards for phenotypic data sets makes it difficult to compare different screening approaches. Here we address this problem in a screen designed to identify all genes required for the first two rounds of cell division in the *Caenorhabditis elegans* embryo. We used RNA-mediated interference to target 98% of all genes predicted in the *C. elegans* genome in combination with differential interference contrast time-lapse microscopy. Through systematic annotation of the resulting movies, we developed a phenotypic profiling system, which shows high correlation with cellular processes and biochemical pathways, thus enabling us to predict new functions for previously uncharacterized genes.**

The concept of 'systems biology' promotes a broad approach to understanding complex biological processes by examining the interplay between molecular assemblies and networks, rather than focusing on individual molecules. Nevertheless, to make this approach possible it is necessary to know which molecules, among the thousands of predicted gene products, actually make up these assemblies and how they might contribute to functionality within the networks. Although achieving this goal is now facilitated by the advent of functional genomics, a key challenge of these approaches, beyond the efficient exploitation of sequence data, is to bridge the gap between large-scale analyses and more in-depth studies by individual laboratories.

The discovery of RNA-mediated interference (RNAi) as a method for targeted gene silencing introduces the best functional genomics tool available for metazoan organisms so far, providing a direct causal link between genes and cellular functions. The basic method harnesses an endogenous gene regulation pathway by which exogenous double-stranded RNA molecules (dsRNAs), after their introduction into cells, are processed to generate a pool of short interfering RNAs (siRNAs) that in turn direct the catalytic degradation of complementary 'target' mRNAs. We and others have previously established the systematic application of RNAi as a genome-scale screening method of identifying and characterizing genes with functions of interest<sup>1-6</sup>. For monitoring individual pathways, such screens can be run with readouts of relatively 'low content' such as single reporter assays, thus enabling a higher throughput. In comparison, higher-content assays, ideally integrating

both spatial and temporal information, enable more comprehensive analyses. However, the challenge arises to provide systematically annotated information that allows comparison to other screening methods, such as two-hybrid data and expression profiling<sup>7,8</sup>.

The primary goal of the present study was to identify all genes essential for mitotic cell division in a metazoan organism. For a mother cell to divide into two daughters correctly, several coordinated events must take place. These include duplication of the centrosome starting in late G1, replication of chromosomal DNA at S phase, assembly of a mitotic spindle at M phase, segregation of the replicated genome to opposite poles of the mitotic spindle at anaphase, and division of the mother cell into two daughter cells at cytokinesis. Previous work has established the early *C. elegans* embryo as an excellent *in vivo* system for detailed studies of metazoan cell division (reviewed in ref. 9). In particular this system allows time-lapse cytological analyses of the first embryonic cell divisions with excellent spatial and temporal resolution<sup>10</sup>. The early *C. elegans* embryo does not arrest in response to cell cycle checkpoint mechanisms<sup>11</sup>, which normally limit the analysis of disruption phenotypes in most other cellular systems. Furthermore, *C. elegans* is amenable to RNAi experimentation<sup>12</sup> and also offers a fully sequenced genome<sup>13</sup>, thereby making it uniquely suited for genome-scale RNAi screening.

## Genome-wide screening using a high-content assay

Our aim was to test all *C. elegans* genes for their role in the first two rounds of mitotic cell division after fertilization. To this end, we



Table 1 **Phenotypic and functional classification of early embryonic (DIC) phenotypes**

| Phenotype class                          | Phenotype description                                                                                                                        | Total (unknown function) | Biochemical pathway                         | Functional class                            | Examples of genes                                                                                                                                                                                                                                                                      |
|------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|---------------------------------------------|---------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sterility/impaired fertility in F0       | Injected worm produces no/few embryos                                                                                                        | 42 (4)                   | Secretion                                   | Production of oocytes                       | F12F6.6(sec-24.1) C39F7.4(rab-1) F26D10.3(hsp-1) F32D8.6(emo-1) T24B8.1(rpl-32) Y113G7A.3(sec-23) ZC518.2(sec-24.2) F43E2.8(hsp-4) W01A11.2 B0336.2(arf-1) C01F6.3(ccp-31A1)                                                                                                           |
| Osmotic integrity                        | Embryo fills egg shell, lyses upon dissection or during recording                                                                            | 109 (15)                 | General cell maintenance                    | Membrane stability, egg shell formation     | C07G2.3a(cct-5) D2045.1(atx-2) F53G12.1(rab-11.1) F57B10.10(dad-1),T25G3.2(chi-1) T20H4.3(prs-1) T23D8.4(eif-3.C)Y47G6A.12(sep-1) Y54E10BL.6(mek-2) Y56A3A.20(ccf-1) Y62E10A.15 Y71F9AL.13a(rpl-1) Y71F9B.4(snr-7) Y76A2B.1(pod-1) Y77E11A.13a(npp-20) Y116A8C.42(snr-1)ZK675.1(ptc-1) |
| Polar body extrusion                     | Unextruded or resorbed polar body(ies) leading to extra PN's in P <sub>0</sub> and/or extra karyomeres in AB/P <sub>1</sub>                  | 12 (2)                   | Defective meiosis, kinesin-like proteins    | Meiosis chromosome segregation              | C06G3.2(klp-18) F52H2.7 F57B10.12(mei-2) M01E11.6(klp-15) Y54E10A.9a(vbh-1) C41G7.2(klp-16) R06A4.4a(imb-2)                                                                                                                                                                            |
| Passage through meiosis                  | Male and female PN's not visible; embryo often fills egg shell completely                                                                    | 47 (4)                   | Cell cycle progression, protein destruction | Meiotic cell cycle progression              | ZK945.2(pas-7) F25B5.4a(ubq-1) C47B2.4(pbs-2) C02F5.9(pbs-6) C30C11.2(rpn-3) C36B1.4(pas-4) F35G12.9(apc-11) K06A5.7(cdc-25.1) T05G5.3(cdk-1) ZK177.6(fzy-1)                                                                                                                           |
| Entry into interphase                    | Embryos spend longer than normal when entering first interphase                                                                              | 4 (0)                    | Cell cycle progression                      | Meiotic cell cycle progression              | C09G4.3(dom-6) F38H4.9 F48E8.5 ZK520.4a(cul-2)                                                                                                                                                                                                                                         |
| Cortical dynamics                        | Little/no cortical ruffling or pseudocleavage furrow, or excessive cortical activity at two-cell stage                                       | 19 (6)                   | Rho, transcription                          | Cortical structure                          | Y51H4A.3(rho-1) Y71F9AL.16(arx-1)                                                                                                                                                                                                                                                      |
| Pronuclear/nuclear appearance            | PN and nuclei are small or missing altogether, often accompanied by spindle defects                                                          | 28 (4)                   | Nuclear pore complex, import machinery      | Reformation of nuclei after meiosis/mitosis | T01G9.4(npp-2) C26D10.1(ran-3) F28B3.8(imb-1) W04D2.1a(atn-1) F26B1.3(lma-2)                                                                                                                                                                                                           |
| Centrosome attachment                    | Centrosomes detach from the male PN                                                                                                          | 3 (2)                    | –                                           | Centrosome attachment                       | F23F12.2 F57B1.2 ZK546.1                                                                                                                                                                                                                                                               |
| Pronuclear migration                     | Lack of male pronuclear migration, female pronuclear migration variable, sometimes multiple female PN's, no or small spindle                 | 38 (3)                   | Dynein–dynactin, microtubule                | Microtubule function                        | F22B5.7(zyg-9) C47B2.3a(tba-2) F13D12.7(gpb-1) C36E8.5(tbb-2) K08E7.3(jet-99) R151.9 T03F6.5(lis-1) T16G12.1 T21E12.4(dhc-1) Y54E2A.3(tac-1) Y108G3AL.1(cul-3) F49E11.1a(mbk-2) F56A3.4(spd-5)                                                                                         |
| Spindle assembly                         | Spindle bipolarity is not established                                                                                                        | 21 (6)                   | Aurora A, centrioles                        | Microtubule function                        | K07C11.2(air-1) F59E12.2(zyg-1) F10E9.8(sas-4)                                                                                                                                                                                                                                         |
| Spindle elongation/integrity             | Bipolar spindle shows clear elongation defect                                                                                                |                          |                                             |                                             | F36A4.7(ama-1) F44A2.3 K11D9.1a(klp-7)                                                                                                                                                                                                                                                 |
| Sister chromatid separation (cross-eyed) | Daughter nuclei are deformed and stay close to central cortex, cytokinesis defects                                                           | 64 (7)                   | Cohesion, kinetochore replication, histone  | Chromosome function                         | C25D7.6(mcm-3) F18E2.3(scc-3) W03D2.4(pcn-1) B0035.8(his-48) B0207.4(air-2) C02F5.1(knl-1) C43E11.10(cdc-6) K12D12.2(npp-3) T06E6.2a(cyb-3) T27F2.3(bir-1) Y54E10A.15a(cdt-1) R06F6.1(cdl-1)                                                                                           |
| Nuclear appearance                       | PN's are normal but nuclei are completely missing or significantly smaller than normal; often accompanied by spindle and cytokinesis defects | 5 (0)                    | Replication, histone                        | Chromosome function                         | F10C2.4 F55G1.11(his-60)Y39G10AR.14(mcm-4)                                                                                                                                                                                                                                             |
| Chromosome segregation (karyomeres)      | Karyomeres in AB and/or P <sub>1</sub> , often accompanied by weak/thin wobbly spindle                                                       | 23 (2)                   | Cyclin, histone                             | Chromosome segregation                      | Y43F4B.6(klp-19) ZC168.4(cyb-1) K06A5.4 T23H2.1(npp-12)                                                                                                                                                                                                                                |
| Cytokinesis                              | Cytokinesis defects in the first and/or second stages of cell division                                                                       | 15 (1)                   | Actin/myosin                                | Cytokinesis                                 | M03F4.2a(act-4) F11H8.4a(cyk-1) C56G7.1(mlc-4) K08E3.6(cyk-4) M03D4.1a(zen-4) Y18D10A.20(pfn-1)                                                                                                                                                                                        |
| Asymmetry of division                    | Symmetric (PAR-like) divisions or excessive posterior displacement (zyg-8 like phenotypes)                                                   | 12 (1)                   | PAR                                         | Polarity                                    | W08F4.8 F09E5.1(pk-3) M117.2(par-5) C38C10.4(gpr-2) F20G4.3(nmy-2) F22B7.13(gpr-1) F58B6.3a(par-2) T25C8.2(act-5) T26E3.3(par-6)                                                                                                                                                       |
| Pace of P-lineage                        | More than 5 min between AB and P <sub>1</sub> divisions                                                                                      | 14 (3)                   | DNA replication                             | DNA damage checkpoint                       | F31E3.3(rfc-4) F33H2.5 R01H10.1(div-1) Y41C4A.10(elt-1) ZK637.8a(unc-32) ZK742.1a(imb-4)                                                                                                                                                                                               |

**Table 1** – continued

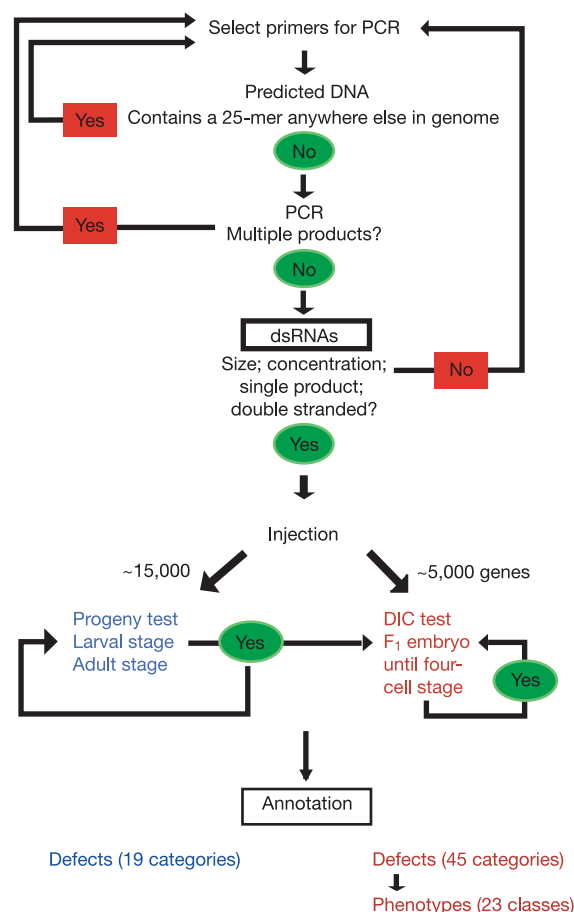
| Phenotype class                          | Phenotype description                                                                                              | Total (unknown function) | Biochemical pathway                                                                     | Functional class       | Examples of genes                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|------------------------------------------|--------------------------------------------------------------------------------------------------------------------|--------------------------|-----------------------------------------------------------------------------------------|------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| General pace of development              | More than 30 min from PN meeting to furrow initiation in AB                                                        | 49 (11)                  | ATP generation                                                                          | Mitochondrial function | Y69A2AR.18a F58F12.1 W10D9.5 T06D8.8(rpn-9) R53.4 T27F7.1 F44G4.2 R06F6.2 F01F1.12a F23B12.5 R11A8.6(irs-1) C37H5.8(hsp-6) T23D8.1(mom-5) C06H2.1 C15H9.10 C34E10.6(atp-2) C53A5.1 C56C10.8 D2030.4 D2085.3 E04A4.5 F23H11.5 F29B9.11 F35G12.2 F35G12.10(asb-1) F36A2.7 F42A6.6 F54H12.1a(aco-2) F56D2.1 F59C6.5 H06H21.3 H28O16.1 K07A12.5 T05H4.12 T07C4.7(mev-1) T09B4.9 T20G5.2 T20H4.5 T22B11.5 W10D5.2 Y37E3.9 K04G7.4a C25H3.9a T05H10.6a Y82E9BR.3a T27E9.1a T10E9.7a W02D9.1(pri-2) C41G7.3                                                                                                              |
| Severe pleiotropic defects               | Often multiple female PN's, aberrant cytoplasmic texture, drop in overall pace of development, osmotic sensitivity | 108 (7)                  | Translation machinery, ribosomal proteins, tRNA synthase, translation initiation factor | Protein synthesis      | F28C6.7a(rpl-26)T05C12.7(cct-1) C01H6.5a(nhr-23) ZC434.5(ers-2) Y41E3.4(qrs-5) F22B5.2(eif-3.G) F22B5.9(frs-2) F56E10.4(rps-27) C47E12.1(srs-2) F28H1.3(ars-2) C37H5.6a C49H3.11(rps-2) W02F12.5 B0041.4(rpl-4) B0250.7 B0336.10(rpl-23) B0348.6a(life-3) B0393.1(rps-0) B0393.8 B0412.4(rps-29) B0412.5 B0464.1(drs-1) B0511.10(eif-3.E) C04F12.4(rpl-14) C09D4.5(rpl-19) C09H10.2 C14B9.7(rpl-21) C16A3.9(rps-13) C52E4.3(snr-4) C53H9.1(rpl-27) F22D6.3(nrs-1) F26F4.10(rrt-1) R08D7.3(eif-3.D) T05G5.10(iff-1) Y87G2A.5(vrs-2) ZC434.2(rps-7) ZK652.4(rpl-35) ZK792.2(inx-8) H06I04.4a(ubl-1) K01C8.10(cct-4) |
| Integrity of membrane-bounded organelles | Sparse or enlarged yolk granules                                                                                   | 13 (2)                   | Secretion/endocytosis                                                                   | Organelle maintenance  | W03C9.3(rab-7) C10C6.6 Y53C12A.1(wee-1.3) C03C10.1(kin-19) C30F8.2 C45G9.5 F18C12.2a(rme-8) H15N14.2(nsf-1) T07C4.8(ced-9) T20G5.1 W08E3.1(snr-2) Y37D8A.10 T23G7.4(sec-5)                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Egg size                                 | Egg is larger or smaller than normal                                                                               | 5 (2)                    | Importins                                                                               | –                      | F32E10.4(ima-3) C53D5.6(lmb-3) F21H12.2 R119.4(pqn-59) C27D9.1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Aberrant cytoplasmic structures          | Areas devoid of yolk granules throughout the embryo                                                                | 9 (6)                    | –                                                                                       | –                      | C56C10.3 K04D7.1 C07G1.5(pqn-9) C09G12.9 CD4.4 E01B7.1 F23C8.6 T17E9.2a T20B12.1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Complex phenotype                        | Complex combination of defects that does not match other class definitions                                         | 21 (0)                   | –                                                                                       | –                      | C08B11.1(zyg-11) C14B9.4a(plk-1) C18E9.6 ZC404.8(spn-4) C39B5.2 C47G2.5 F07A5.1a(inx-14) F25H2.4 F26H9.6(rab-5) F43C1.2a(mpk-1) F45H11.2(ned-8) H38K22.2a M7.1(let-70) Y18D10A.17 Y37B11A.3 Y56A3A.18 Y71H2B.3 Y75B8A.30(pph-4.1) Y82E9BR.15(elc-1) F14B8.1a(nhx-4) B0334.11a(oc-3)                                                                                                                                                                                                                                                                                                                               |
| All early embryonic phenotypes           | Phenotypes scored in DIC test                                                                                      | 661 (94)                 | Cell division/general metabolism                                                        | –                      | –                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |

Representative examples are shown for genes in 23 phenotype classes, and their association with specific biochemical pathways and functions. AB and P<sub>1</sub>, anterior and posterior blastomere, respectively, of the two-cell embryo; PAR genes, genes associated with polarity of the embryo; PN, pronucleus.

combined genome-wide RNAi screening with a time-lapse microscopy readout. As a basic method we injected dsRNA into young adult worms and examined the phenotype of the resulting F<sub>1</sub> progeny<sup>2</sup>. To design a genome-wide set of inhibitory dsRNAs, each targeting only one gene, an earlier algorithm used to generate dsRNAs for chromosome III (ref. 2) was updated to optimize further the targeting specificity of dsRNAs by taking into account the fact that the average length of siRNAs in *C. elegans* is 25 base pairs (bp) (ref. 14). Thus, whenever possible, dsRNAs were designed to avoid contiguous siRNA-sized (25 bp) stretches of complementarity to off-target transcriptome sequences. We were able to design dsRNAs to 16,465 (84.2%) genes with this stringency (according to Wormbase 100; see Methods). For an additional 14% of genes, dsRNAs were designed by allowing one or more 25-mers to be present in more than one gene. By this approach, 20,326 dsRNAs targeting 19,075 (98%) of the genes in *C. elegans* were designed, produced and screened. Continuous changes in gene annotation have meant that gene coverage by the designed dsRNAs was regularly updated during the course of the project.

The primary phenotypic analysis was performed with time-lapse differential interference contrast (DIC) microscopy of F<sub>1</sub> embryos from about 20 min after fertilization until the four-cell stage, 35 min later ('DIC test'; see also ref. 2). In addition, F<sub>1</sub> broods from individual injected parents were examined to determine the viability

of the embryo as well as gross phenotypic defects at larval and adult stages ('progeny test') (Fig. 1). We first defined a set of 2,093 genes that, on the basis of their expression pattern and various bioinformatics criteria, were expected to have a higher likelihood of exhibiting functions in early embryogenesis. Each of these genes was tested individually by both the DIC and progeny tests, as described above. Results from this initial 2,093-gene set and from our earlier chromosome III study<sup>2</sup> showed that virtually all genes exhibiting early embryonic phenotypes detectable by the DIC test also show embryonic lethality in the progeny test. Therefore, dsRNAs targeting the remaining genes, which were expected to yield a significantly lower hit rate, were initially injected in pairs, and analysed by the less laborious progeny test. All dsRNAs from embryonic lethal pairs were then reinjected individually and analysed with both tests. Each incidence of a detectable phenotype was confirmed by at least two independent experiments (see Methods). During the course of the study, more than 40,000 time-lapse DIC recordings from over 19,000 individual experiments were acquired and annotated. Necessary for the interpretability of the final data set was the strict standardization of all procedures throughout this large-scale endeavour, from data collection to data annotation and analysis. For this purpose, data collection was automated by a customized laboratory information management system developed specifically for this project. The resulting data set is available online at <http://www.worm.mpi-cbg.de/phenobank2/>.



**Figure 1** Flowchart for the screening process.

## 661 genes required for early embryogenesis

Of the 20,326 dsRNAs tested so far, 1,995 dsRNAs targeting 1,668 genes, corresponding to 8.5% of all predicted genes in *C. elegans*, were found to elicit detectable RNAi-induced phenotypic alterations in at least two distinct experiments (Fig. 2).

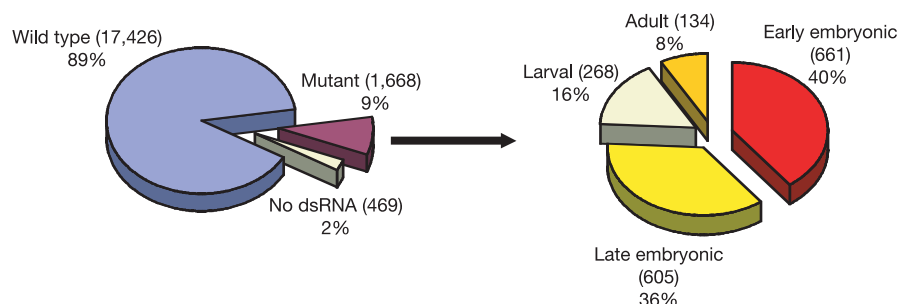
We focused our attention particularly on the 661 genes that reproducibly exhibited altered phenotypes in the early embryo (Fig. 2 and Table 1, and Supplementary Table 1). How comprehensive was our screen in detecting those loss-of-function DIC phenotypes? Because the interpretation of negative results in RNAi screens is always problematic, we used two different methods to address this

issue. First, we compared the results from our RNAi screen with previously published conventional genetic analyses (Supplementary Table 2). Of the 65 genes previously shown by mutagenesis to be required during the first cell division, we obtained corresponding RNAi phenotypes for 62 or 95% of these genes. Retesting the three missed genes with the same batch of dsRNAs increased this success rate to 64 (98%) of the 65 genes, thus ascribing most of the initial failures to technical variability. These data would suggest that over 90% of genes required for the first cell division were detectable by RNAi under our screening conditions.

However, mutagenesis-based screens are typically limited in their coverage because of their tendency to focus on genes that yield robust and easily identifiable mutant phenotypes. We therefore also compared our data with two other RNAi-based screens<sup>7,15</sup> that similarly documented early embryonic phenotypes by using DIC recordings in *C. elegans* (Supplementary Table 3). We note that a significant proportion of those genes found by others to be DIC positive but annotated by us as DIC negative did, in fact, exhibit detectable DIC defects, although with insufficient reproducibility to match our criteria (see online database for detailed documentation). Furthermore, 75% of these ‘missed DIC genes’ also displayed altered phenotypes in our progeny tests. These analyses suggest a DIC detection accuracy of between 75% and 90% for our screen.

## From phenotypic annotation to gene function

The biggest challenge of this screen was to convert the data from more than 40,000 recordings into standardized, searchable text annotations. The complexity of the DIC recordings rendered automated analyses unfeasible. Thus, to minimize the inherent variability of visual evaluations, all data sets were annotated by at least two scientists, and importantly, one person (B.S.) finalized the analysis of all recordings, giving the annotation a unity lacking when individual genes are analysed in separate, individual laboratories. We based our annotation of DIC data on a descriptive classification of all deviations from the wild type. Every DIC recording was scored for 45 distinct defect categories (Table 2). Each experiment comprised five DIC recordings of single embryos originating from five different injected worms. Reproducibility was documented as the percentage of DIC recordings exhibiting a particular defect within a given experiment (see Methods). With this strategy we achieved a high resolution of phenotypic annotation, by which combinations of defects defined gene-specific phenotypic signatures<sup>7</sup>. After identifying patterns of highly reproducible ‘core’ defects, we manually grouped the 661 DIC-positive genes according to their phenotypic signatures, thereby defining 23 phenotype classes (Figs 3 and 4 and Table 1). The defect pattern for each gene can also be presented by assigning a single-digit score of reproducibility to each of the 45 defect categories. These scores are graphically represented as distinct



**Figure 2** Distribution of phenotypes identified in the *C. elegans* genome-wide RNAi screen. A total of 19,075 genes were targeted by dsRNAs. Progeny tests were performed to assess late embryonic, larval and adult phenotypes. Early embryonic phenotypes were

assessed by filming the first two rounds of cell divisions with time-lapse DIC microscopy. Each incidence of a phenotype was confirmed by two independent experiments.



Table 2 Defect categories and associated scoring criteria

| No. | Defect category                                    | Scoring criteria                                                                                                                                                               |
|-----|----------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1   | Egg shape/size                                     | Eggs small (less than 70% of wild type), large (more than 130% of wild type) or irregularly shaped                                                                             |
| 2   | Osmotic sensitivity                                | Swelling of the embryo to fill the egg shell and/or irregular granule movements                                                                                                |
| 3   | P <sub>0</sub> cortical ruffling                   | Excessive or no early cortical ruffling (wild type: wave of membrane ruffling from posterior to anterior resulting in pseudo-cleavage furrow)                                  |
| 4   | P <sub>0</sub> cytoplasmic flows                   | Lack of granular flows towards the male PN (wild type: directional flow of yolk granules towards the male PN)                                                                  |
| 5   | P <sub>0</sub> pseudo-cleavage furrow              | No or excessive furrowing (wild type: covering 20–30% the width of the embryo)                                                                                                 |
| 6   | P <sub>0</sub> pronuclei – size/shape              | Size more than 30% smaller or larger than wild type (wild type: diameter approx. 25% the width of the embryo at onset of PN migration), or irregular shape of PNs              |
| 7   | P <sub>0</sub> pronuclei – number                  | Numbers of PNs above or below 2 (wild type: one female, one male PN)                                                                                                           |
| 8   | P <sub>0</sub> pronuclear migration (female)       | No or little migration of female PN towards the male PN (wild type: directional movement of the female towards the male PN)                                                    |
| 9   | P <sub>0</sub> pronuclear migration (male)         | No migration of male PN from posterior cortex towards centre of the embryo                                                                                                     |
| 10  | P <sub>0</sub> pronuclear meeting                  | Position near the cortex or centrally (wild type: near the centre of the posterior half of the embryo)                                                                         |
| 11  | P <sub>0</sub> pronuclear centration               | Lack of centration                                                                                                                                                             |
| 12  | P <sub>0</sub> pronuclear rotation                 | Rotation takes place after PN envelope breakdown (wild type: approx. 2–3 min before Pronuclear Envelope Breakdown)                                                             |
| 13  | P <sub>0</sub> pronuclear envelope breakdown       | Lag time between PN meeting and PN envelope breakdown exceeds 8 min (wild type: 4–5 min)                                                                                       |
| 14  | P <sub>0</sub> spindle assembly – bipolarity       | Lack of two visible poles, that is, two regularly shaped sites of yolk granule exclusion                                                                                       |
| 15  | P <sub>0</sub> spindle integrity                   | Irregular length (wild type: 25–30% the length of the embryo during metaphase), or thickness (wild type: 20–25% the width of the embryo at metaphase), or lack of rigidity     |
| 16  | P <sub>0</sub> spindle elongation                  | Spindle is shorter or longer than 50–60% the length of the embryo at telophase                                                                                                 |
| 17  | P <sub>0</sub> spindle rocking                     | No or excessive spindle rocking                                                                                                                                                |
| 18  | P <sub>0</sub> spindle positioning – asymmetry     | Aberrant spindle positioning at telophase (wild type: along the longitudinal axis with the posterior pole shifted posteriorly approx. 10–15%)                                  |
| 19  | P <sub>0</sub> spindle poles                       | Irregular shape, in particular lack of flattening of posterior pole in telophase                                                                                               |
| 20  | P <sub>0</sub> cytokinesis – furrow specification  | Fewer or more than two sites of furrowing and/or aberrant positioning (wild type: two sites, intersecting the long axis by a ratio of approx. 3:2)                             |
| 21  | P <sub>0</sub> cytokinesis – furrow ingression     | Little or no ingression, or uneven ingression from the two sites                                                                                                               |
| 22  | P <sub>0</sub> cytokinesis – completion/stability  | Regression of the furrow                                                                                                                                                       |
| 23  | P <sub>1</sub> /AB nuclear separation – cross-eyed | Reforming daughter nuclei stay closely attached to the central cortex                                                                                                          |
| 24  | AB nuclear migration                               | Time for centration of AB nucleus exceeds 7–8 min, AB nucleus migrates towards the cortex before centration (wild type: AB nucleus usually centres directly after cytokinesis) |
| 25  | P <sub>1</sub> /AB cortical activity               | Excessive membrane ruffling and blebbing                                                                                                                                       |
| 26  | P <sub>1</sub> /AB nuclei – number                 | Numbers of nuclei in daughter cells below or above 1                                                                                                                           |
| 27  | P <sub>1</sub> /AB nuclei – size/shape             | Size >30% smaller or larger than WT (diameter approx. 25% of AB blastomere), or irregular shape of nuclei                                                                      |
| 28  | P <sub>1</sub> nuclear migration/rotation          | Lack of migration of P1 nucleus towards posterior cortex, lack of rotation of P1 spindle before nuclear envelope breakdown                                                     |
| 29  | P <sub>1</sub> /AB spindle assembly                | Aberrant bipolarity or length or thickness                                                                                                                                     |
| 30  | AB spindle orientation                             | Rotation of AB spindle (wild type: AB spindle keeps orientation, whereas P1 spindle rotates)                                                                                   |
| 31  | P <sub>1</sub> /AB asynchrony of divisions         | Delay between AB and P1 cleavage furrow initiation is either shorter than 2 min or exceeds 5 min (wild type: approx. 2–3 min)                                                  |
| 32  | P <sub>1</sub> /AB cytokinesis                     | Aberrant furrow initiation or ingression or completion                                                                                                                         |
| 33  | Four-cell stage cross-eyed                         | Reforming daughter nuclei stay closely attached to the central cortex                                                                                                          |
| 34  | Four-cell stage nuclei – size/shape                | Irregular size and/or shape of nuclei in daughter cells                                                                                                                        |
| 35  | Four-cell stage configuration                      | PAR-like configurations of blastomeres                                                                                                                                         |
| 36  | Tetrapolar spindle                                 | Four poles, visible by exclusion of yolk granules                                                                                                                              |
| 37  | Yolk granules – density                            | Reduction more than 30%                                                                                                                                                        |
| 38  | Yolk granules – size                               | Aberrant size of individual or all granules                                                                                                                                    |
| 39  | Areas devoid of yolk granules                      | Aberrant cytoplasmic structures excluding yolk granules                                                                                                                        |
| 40  | Polar bodies                                       | Aberrant number (fewer or more than two) or size (matches or exceeds size of early PNs) or internalization of polar bodies                                                     |
| 41  | Unclear – multinucleate                            | Aberrant numbers of nuclei                                                                                                                                                     |
| 42  | Overall pace of events                             | Time span between pronuclear meeting and initiation of AB cleavage furrow exceeded 30 min (wild type: 18–22 min)                                                               |
| 43  | Number/age of embryos                              | Absence or reduction of one-cell and two-cell stage embryos, suggesting reduced fertility of the parent worm                                                                   |
| 44  | Meiotic arrest                                     | No visible PNs, few or no cytoplasmic events, embryo often fills egg shell                                                                                                     |
| 45  | Other                                              | Any other observation deferring from wild-type events                                                                                                                          |
| 46  | Inadequate test                                    | Technical inadequacy, (focusing, coverage of recordings, etc.)                                                                                                                 |

AB and P<sub>1</sub>, anterior and posterior blastomere; PAR genes, genes involved in polarity of the embryo; PN, pronucleus.

colours in Fig. 3. This type of digital representation will help to facilitate further classification efforts, for example through the application of clustering algorithms.

To extract new insights into gene function from this data set, we first determined whether phenotype classes correspond to specific biochemical processes by searching each class for genes of known

function. Using this approach, we found that 16 of the 23 defined phenotype classes indeed implicated specific biochemical pathways (Table 1). For example, 25 of 49 genes in the ‘pace of development’ class are known components required for ATP metabolism<sup>2</sup>. Similarly, 17 of 28 genes in the ‘pronuclear/nuclear appearance’ class are known to be involved in nuclear transport or nuclear pore

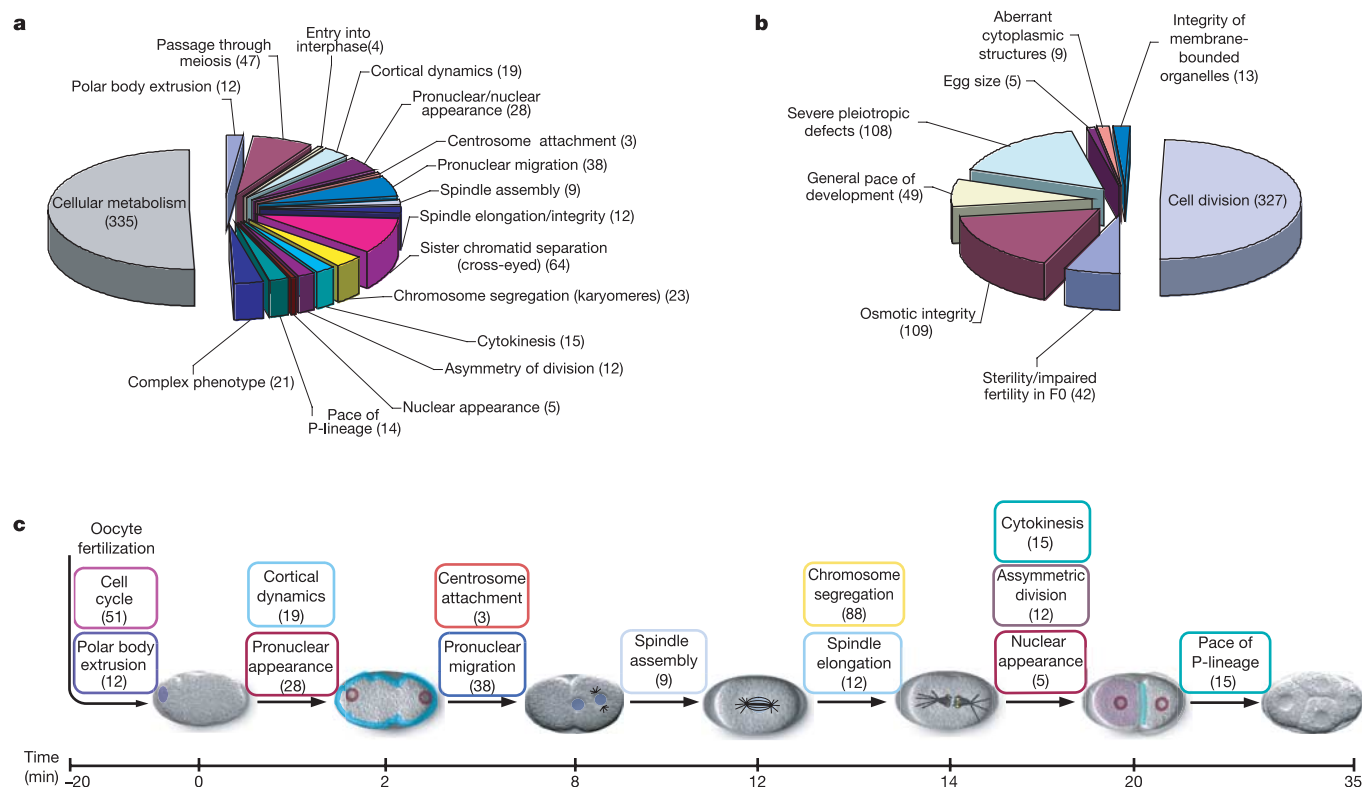
Table 3 Components of protein complexes give rise to distinct common phenotypes

| Protein complex                    | Total no. of genes | Genes scored |     | Assigned phenotype class          |
|------------------------------------|--------------------|--------------|-----|-----------------------------------|
|                                    |                    | No.          | %   |                                   |
| Anaphase-promoting complex         | 9                  | 6            | 67  | 100% passage through meiosis      |
| Minichromosome maintenance         | 6                  | 6            | 100 | 83% sister chromatid separation   |
| Nuclear pore protein               | 20                 | 14           | 70  | 79% pronuclear/nuclear appearance |
| Ribosome protein small subunit     | 26                 | 24           | 92  | 79% severe pleiotropic defects    |
| Ribosome protein large subunit     | 42                 | 38           | 90  |                                   |
| α proteasome subunit 20S particle  | 7                  | 7            | 100 | 93% passage through meiosis       |
| β proteasome subunit 20S particle  | 7                  | 7            | 100 |                                   |
| Non-ATPase subunit of 19S particle | 11                 | 9            | 82  | 80% passage through meiosis       |
| ATPase subunit 19S particle        | 6                  | 6            | 100 |                                   |
| F <sub>1</sub> -ATPase             | 10                 | 9            | 90  | 55% general pace of development   |
| Dynein–dynactin                    | 12                 | 11           | 92  | 100% pronuclear migration         |

Analyses were performed of the total number of genes in a complex, the number of genes associated with a phenotype in the DIC assay, and the predominantly assigned phenotype class.



experiment for each of the 661 genes showing reproducible phenotypes. The colour code resembles the reproducibility of the defect (scores out of 5): red, 5; orange, 4; yellow, 3; green, 2; blue, 1; black, 0. One example gene with associated scores for defects is shown at the top.



**Figure 4** Genes in phenotype classes are associated with distinctive biochemical pathways and key events in the first cell divisions of *C. elegans* embryos. **a**, Phenotype classes of genes required for the process of cell division. **b**, Phenotype classes of genes required for cellular metabolism. **c**, Phenotype classes marking key biological events in the first round of cell division. Numbers of genes associated with each class, according to

the present study, are given in parentheses. The classes 'passage through meiosis' and 'entry into interphase' are summarized as 'cell cycle', and the classes 'sister chromatid separation' and 'chromosome segregation' are summarized as 'chromosome segregation'.

assembly<sup>16</sup>, and 16 of 38 genes in the 'pronuclear migration' class are components required for microtubule function, including members of the dynein–dynactin complex<sup>17</sup>, tubulins and their cofactors. To confirm the accuracy of the phenotypic classification, we also examined how many genes that are thought to be part of the same protein complex gave rise to the same phenotype (Table 3). This revealed that 11 of 12 identified genes of the dynein–dynactin complex were indeed contained within the pronuclear migration class, 79% (54 of 68) of known ribosomal components were classified as 'severe pleiotropic defects'<sup>2</sup>, and 93% (13 of 14) of known proteasome subunits were contained within the 'passage through meiosis' class<sup>18</sup>.

Thus, the strength of our combined phenotypic and bioinformatics analysis gives us considerable predictive power with respect to the function of previously uncharacterized genes. For example, bioinformatic analyses of genes in the respective phenotype classes allows us to propose 7 novel genes for chromosome function, 4 for nuclear envelope assembly and 4 for cell cycle progression, as well as 11 previously undescribed genes required for ATP metabolism and 7 for protein synthesis (Table 1). On the basis of this analysis, we found that 14% (94 of 661) of genes shown in this study to be required for the first cell divisions had previously not been assigned a function in any organism. Of these, about half had homologues in human, mouse, fly, yeast or *Arabidopsis* (see Supplementary Table 1). It should be noted that although DIC optics allowed us a very broad and detailed view of cellular events, more subtle defects in certain processes such as chromosome segregation cannot be seen with this technique. Nevertheless, one expects that such defects would most probably have caused detectable levels of embryonic lethality scored in our progeny testing. Time-lapse analyses of

fluorescently tagged markers in living embryos are now underway to provide further insights into these processes.

## Conclusion

Our data suggest that between 650 and 800 individual genes are required for early embryogenesis under laboratory conditions. Roughly half of these genes are involved in cell division processes such as chromosome segregation or cytokinesis, and the other half are needed for cell maintenance processes such as translation and mitochondrial function (Fig. 4 and Table 1).

In view of gene deletion studies that have identified more than 1,100 genes as being essential for *Saccharomyces cerevisiae*<sup>19</sup>, our present tally for *C. elegans* early embryos might appear surprisingly low. One possible explanation could be that the use of RNAi inherently underestimates the number of genes required for this system. However, our results indicate that the RNAi approach used here proved at least equal to the conventional ethylmethane sulphonate mutagenesis-based studies in identifying loss-of-function phenotypes.

Another possibility is that, unlike the embryo growing within the protective environment of the egg shell, somatic divisions such as those in *S. cerevisiae* or adult metazoan cells might be more complex because they must respond to external stimuli and insults arising from their more complex surroundings. Such a response requires robust signal transduction processes regulating all aspects of cell division. Early embryonic divisions follow a programmed course of division that might prove much simpler. A third possibility is that *C. elegans*, as a member of the nematode family with a particularly short development time, has at some stage simplified its cell division, perhaps owing to evolutionary pressure. Comparison



with RNAi screens in other organisms should help in distinguishing between these different ideas.

The present study illustrates the power of combining genome-scale RNAi with high-content assays to bridge the gap between systems-level biology and traditional single-gene-based analyses of function. Beyond attributing genes to various cell division processes, we were able to put our analysis into context by documenting a wider range of basic cellular functions, therefore creating a broader data set that permits exploitation by researchers from diverse fields. In particular, the ability to represent complex phenotypic data as digital signatures will be essential for integrating and comparing new data sets from disparate experimental sources. □

## Methods

### Generation of oligonucleotide primer pairs

Polymerase chain reaction primer sequences were designed by using the customized algorithm described in ref. 2, updated to avoid contiguous siRNA-sized (25-bp) stretches of complementarity to off-target transcriptome sequences. All primer sequences are available in the online database (<http://www.worm.mpi-cbg.de/phenobank2/>).

### Generation of dsRNA

Synthesis of dsRNAs *in vitro* was performed as described in ref. 2.

### Experimental set-up and phenotype readout criteria

dsRNAs were injected bilaterally into gonads of N2 hermaphrodites, maintained in accordance with standard procedures<sup>20</sup>. The animals were left at 20 °C for 24 h. Note that, in contrast to our chromosome III study<sup>2</sup>, pairs of dsRNAs were injected only for progeny test experiments.

For the progeny test, three injected animals were transferred to three individual wells of a six-well plate 24 h after injection of the dsRNA, and left at 18 °C. Three days later, the wells were checked for the presence and developmental stage of F<sub>1</sub> larvae (L1 to L4). Three days after that, the plate was inspected again for the presence and overall body morphology of F<sub>1</sub> adults and F<sub>2</sub> progeny.

Each animal generates well over 50 wild-type-looking F<sub>1</sub> progeny (wild type). In general, dsRNAs that caused defects in less than 10% of the progeny were not considered to be associated with a significant phenotype. Animals that produced fewer than 15 progeny were deemed to be sterile or to have fertility problems. Embryonic lethality was determined by counting all F<sub>1</sub> progeny at the stage of eggs and larvae (3 days after being shifted to 18 °C) and subsequently at the adult stage (6 days after being shifted to 18 °C). The number of unhatched eggs was then calculated and expressed as a percentage of the overall F<sub>1</sub> progeny.

Larval phenotypes fell into one of two groups: those in which the progeny at the final check resembled L1 or L2 animals, and those in which the progeny at this check resembled L3 or L4 animals. Other dsRNAs gave rise to a clear morphological defect in larvae or adults, such as Dpy or Unc; some were associated with slower development.

For the time-lapse DIC microscopy assay, all time-lapse experiments were performed under tight temperature control of the room at 21 ± 2 °C. Embryos were removed from the injected animals and recorded as described in ref. 2. A minimum of five embryos from five different worms were filmed, three of them from shortly after fertilization until the four-cell stage, and the remaining two for 5–10 min.

All recordings of early embryos were scored for 45 different criteria (see Table 2). Additional deviations from the wild type were noted when apparent (category ‘other’). This resulted in defect patterns for individual dsRNAs, summarizing all observed defects with their respective reproducibility. All phenotypes were reproduced in at least two independent experiments. Whenever a phenotype was observed in only one or two of five recordings, but could be reproduced with a similar score in an independent experiment, it was termed a phenotype of low reproducibility. In comparison, phenotypes we describe to be of high reproducibility were scoring in four or five of five recordings in each experiment. Note that in most cases we chose to perform confirmation rounds with the same dsRNA, because the data from our chromosome III study<sup>2</sup> showed that a new ‘second-best’ dsRNA design often resulted in a weaker manifestation of the observed phenotype.

Similar defect patterns were manually grouped into 23 phenotype classes. To illustrate this clustering of phenotypic signatures graphically, one representative experiment was chosen for each of the 661 genes associated with a DIC phenotype, and each of the 45 defect categories was assigned a single-digit score according to its reproducibility in this experiment. These scores were then represented graphically by distinct colours (Fig. 3).

Minor deviations that were also occasionally observed in the wild type were scored at the defect level but were not taken into account for the phenotypic classification. These included slight irregularities in size for subpopulations of yolk granules, incomplete

rotation of the pronuclei in P<sub>0</sub>, and a slight delay for nuclear centration in the AB blastomere.

### Bioinformatic analysis

All statistics on genome coverage presented in this study are based on mapping to version 100 of the Wormbase genome annotation, which is the first ‘frozen’ genome annotation for *C. elegans*.

To identify orthologues in other species, Blast search of the relevant *C. elegans* protein was performed with the Blastp sequence-analysis program<sup>21</sup>. The sequence was searched against a collection of model organism genome databases, human, mouse, rat, *Drosophila* and yeast at the National Center of Biotechnology Information. A sequence was identified as an orthologue if it matched with an expected value of less than 10<sup>−5</sup> and was itself the reciprocal blast hit of the *C. elegans* gene. If the sequence match was less than 10<sup>−5</sup> but not a reciprocal hit, the sequence was identified as a homologue. Functional information on individual genes was retrieved from GeneOntology and Locus link.

Received 27 August 2004; accepted 10 January 2005; doi:10.1038/nature03353.

- Lum, L. *et al.* Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells. *Science* **299**, 2039–2045 (2003).
- Gönczy, P. *et al.* Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**, 331–336 (2000).
- Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
- Clemens, J. C. *et al.* Use of double-stranded RNA interference in *Drosophila* cell lines to dissect signal transduction pathways. *Proc. Natl Acad. Sci. USA* **97**, 6499–6503 (2000).
- Kiger, A. *et al.* A functional genomic analysis of cell morphology using RNA interference. *J. Biol.* **2**, 27.1–27.15 (2003).
- Maeda, I. *et al.* Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. *Curr. Biol.* **11**, 171–176 (2001).
- Piano, F. *et al.* Gene clustering based on RNAi phenotypes of ovary-enriched genes in *C. elegans*. *Curr. Biol.* **12**, 1959–1964 (2002).
- Walhout, A. J. *et al.* Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline. *Curr. Biol.* **12**, 1952–1958 (2002).
- Schneider, S. Q. & Bowerman, B. Cell polarity and the cytoskeleton in the *Caenorhabditis elegans* zygote. *Annu. Rev. Genet.* **37**, 221–249 (2003).
- Gönczy, P. *et al.* Dissection of cell division processes in the one cell stage *Caenorhabditis elegans* embryo by mutational analysis. *J. Cell Biol.* **144**, 927–946 (1999).
- Strome, S. & Wood, W. B. Generation of asymmetry and segregation of germ-line granules in early *C. elegans* embryos. *Cell* **35**, 15–25 (1983).
- Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
- The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
- Parrish, S. *et al.* Functional anatomy of a dsRNA trigger: differential requirement for the two trigger strands in RNA interference. *Mol. Cell* **6**, 1077–1087 (2000).
- Zipperlen, P. *et al.* Roles for 147 embryonic lethal genes on *C. elegans* chromosome I identified by RNA interference and video microscopy. *EMBO J.* **20**, 3984–3992 (2001).
- Askjaer, P. *et al.* Ran GTPase cycle and importins alpha and beta are essential for spindle formation and nuclear envelope assembly in living *Caenorhabditis elegans* embryos. *Mol. Biol. Cell* **13**, 4355–4370 (2002).
- Gönczy, P. *et al.* Cytoplasmic dynein is required for distinct aspects of MTOC positioning, including centrosome separation, in the one cell stage *Caenorhabditis elegans* embryo. *J. Cell Biol.* **147**, 135–150 (1999).
- Davy, A. *et al.* A protein–protein interaction map of the *Caenorhabditis elegans* 26S proteasome. *EMBO Rep.* **2**, 821–828 (2001).
- Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
- Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Srayko, C. Cowan, L. Pelletier and W. B. Wood for critical reading of the manuscript, and M. Volkmer and S. Schloissnig for their help on the completion of the online Phenobank database. This work was supported in part by the German government through a DLR grant. A.C. was supported by the MRC, UK.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to B.S. (soennichsen@cenix-bioscience.com).

# Encoding social signals in the mouse main olfactory bulb

Da Yu Lin<sup>1</sup>, Shao-Zhong Zhang<sup>2</sup>, Eric Block<sup>2</sup> & Lawrence C. Katz<sup>1</sup>

<sup>1</sup>HHMI and Department of Neurobiology, Box 3209, Duke University Medical Center, Durham, North Carolina 27710, USA

<sup>2</sup>Department of Chemistry, University at Albany, State University of New York, Albany, New York 12222, USA

**Mammalian urine releases complex mixtures of volatile compounds that are used in reproduction, territoriality and conspecific recognition. To understand how such complex mixtures are represented in the main olfactory bulb, we analysed the electrophysiological responses of individual mitral cells to volatile compounds in mouse urine. In both males and females, urine volatile compounds evoke robust responses in a small subset of mitral cells. Fractionation of the volatile compounds using gas chromatography showed that out of the hundreds of compounds present, mitral cells are activated by single compounds. One cohort of mitral cells responded exclusively to male urine; these neurons were activated by (methylthio)methanethiol, a potent, previously unknown semiochemical present only in male urine. When added to urine, synthetic (methylthio)methanethiol significantly enhances urine attractiveness to female mice. We conclude that mitral cells represent natural odorant stimuli by acting as selective feature detectors, and that their activation is largely independent of the presence of other components in the olfactory stimulus.**

Much of the social signalling between mammals involves the secretion, detection and interpretation of chemical cues. In addition to non-volatile, species-specific signals (which require physical contact between an animal and an odour source), mammals use numerous other chemical signals that are secreted in urine or other bodily fluids and can be detected at some distance from their source. Ewes, for example, recognize their offspring based on volatiles<sup>1</sup>, newborn rabbit pups locate their mothers' mammary glands using a volatile compound found in milk<sup>2</sup>, and mice can recognize one another based on diffusible signals originating from urine<sup>3</sup>. Whereas non-volatile cues are detected by the accessory olfactory system, volatile social cues are processed by the anatomically separate main olfactory system, which relies on an entirely distinct set of peripheral receptor molecules (see refs 4, 5 for recent reviews).

Mouse urine, like most natural olfactory stimuli, consists of a plethora of distinct volatile compounds, of which only a fraction are known. The presence and level of many urine components vary according to the sex, strain, social status and oestrous cycle of individual mice<sup>6–9</sup>. Although this suggests that urine contains information that could be used to assign identity to conspecific mice, the chemical identity of the salient cues, together with the process by which individual neurons encode this complex mixture, remain unknown. To investigate how the stimulus features of urine volatile compounds are encoded by individual mitral cells, we used a combination of extracellular single-unit recordings and gas chromatographic separations of urine components.

## Urine-responsive mitral cells in the olfactory bulb

Urine activates immediate early gene expression in glomeruli located in the ventral–lateral region of the main olfactory bulb<sup>10,11</sup>. We began searching for urine-responsive mitral cells in this region in female mice and located two clusters, on the ventral and lateral aspects of the bulb (each extending about 300  $\mu$ m along the anterior–posterior axis) in which urine-responsive mitral/tufted cells were present at moderate densities (Supplementary Fig. 1). The relative positions of these two clusters potentially reflects the activation of mirror-selective glomerular maps<sup>12,13</sup>. To assess whether urine-responsive neurons were found elsewhere, we systematically recorded from neurons at 100- $\mu$ m intervals along the anterior–posterior, medial–lateral, and dorsal–ventral axes, sampling roughly two-thirds of the bulb (see Methods). In total,

we recorded from 2,733 mitral/tufted cells in the main olfactory bulbs of 102 sexually mature BALB/c female mice. Outside the responsive clusters, urine-responsive cells (neurons for which firing rates were elevated by >30% during stimulus presentation) were rare (Supplementary Fig. 2). A total of 208 neurons (almost all recorded in the responsive clusters) responded to volatiles derived from diluted urine of at least one mouse strain and sex (Fig. 1). This number substantially overestimates the overall frequency of urine-responsive mitral cells (see Supplementary Fig. 1c), as responsive regions were sampled with many more electrode penetrations. Subsequent recordings from an additional 536 neurons in male mice (11 BALB/c and 7 C57Bl/6 mice), focusing on the ventral–lateral regions, yielded a similar proportion (10%) of urine-responsive neurons ( $n = 53$ ). Post-hoc analyses of a random sample of neurons ( $n = 113$ ) indicated that ~5% of neurons ( $n = 6$ ) classified as unresponsive had changes of >2 standard deviations in their firing rates during stimulus presentations (as expected from a random distribution), whereas all neurons classified as responsive ( $n = 50$ ) had changes of >2 standard deviations (we cannot exclude the possibility that very weakly responsive neurons might be missed using our screening procedure). Thus, despite the chemical diversity of the dilute urine stimulus, the mitral cells responsive to urine in both male and female mice comprise a small proportion of neurons, confined to restricted regions of the main olfactory bulbs.

## Mitral cell responses to fractionated urine

Responses to urine began shortly after the onset of the first inhalation following stimulus presentation, and continued for the duration of the 2-s stimulus. Almost all responses were excitatory (241/261, 92%), and were frequently synchronized to the animal's breathing cycle (Fig. 1a, b). A small fraction of neurons (15/261, 6%) were inhibited by urine volatiles (Fig. 1c), and an even smaller number (5/261, 2%) were excited by some types of urine and inhibited by others.

To distinguish whether the activation of a mitral cell in response to urine reflects the integration of responses to several distinct volatile components, or a selective response to an individual component, we combined solid-phase microextraction (SPME) of urine volatiles with gas chromatography (GC) and single-unit electrophysiological recording (gas chromatography-electrophysiology, or GC-E<sup>14,15</sup>).

SPME is used to concentrate volatiles in the headspace above a volume of urine by dissolving them in a series of coatings that adsorb compounds of low molecular mass on the basis of size and polarity. The total amount of a volatile absorbed by a SPME fibre is linearly proportional to its initial concentration in urine<sup>16,17</sup>. To minimize the bias of SPME fibres towards chemicals with certain molecular weights and polarities, we used a fibre consisting of three phases: Carboxen, divinylbenzene and polymethylsiloxane. Although these are currently the most 'unbiased' fibres available<sup>18–20</sup> they nonetheless concentrate some molecules to a greater extent than others, and might not concentrate some volatile compounds effectively (see Discussion). Despite these potential caveats, this approach allowed us to record the activity of urine-

responsive mitral cells to at least 100 sequentially presented volatile components of urine (Supplementary Fig. 3).

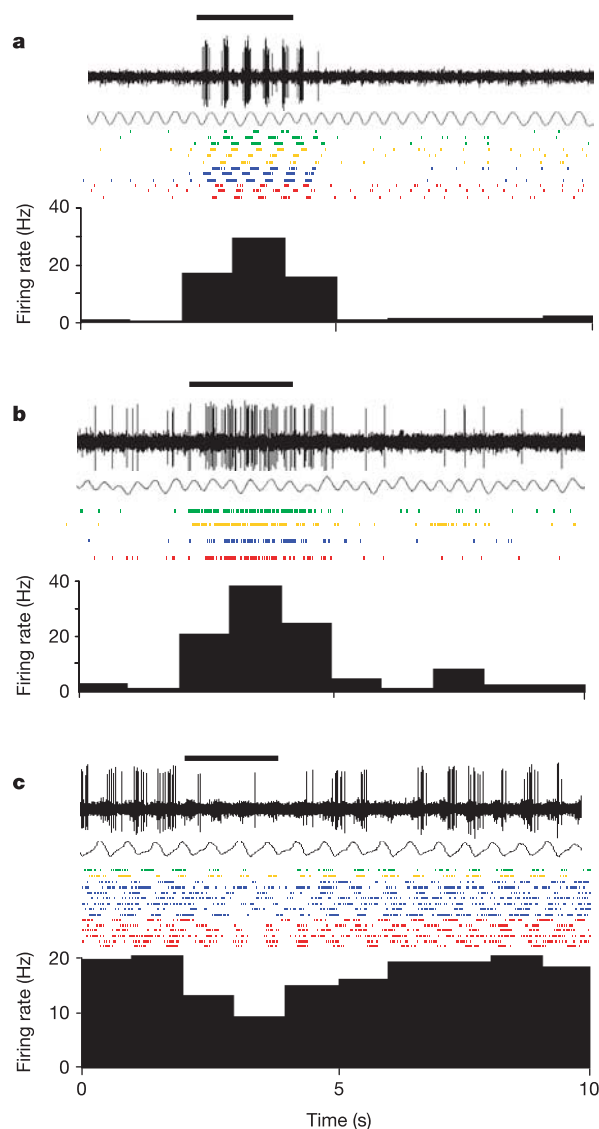
We used GC-E to assess the responses of 104 urine-responsive mitral cells (from 57 BALB/c female mice, 10 BALB/c male mice and 4 C57Bl/6 male mice) to separated volatile components. For 83 neurons, responses were confirmed by at least two independent GC runs. After isolating neurons that responded to urine volatiles (Fig. 2a, d), we continuously recorded action potentials for 20 min while eluting volatiles that had been concentrated by SPME (see Methods). Responses to eluted compounds were unambiguous: an abrupt increase in firing rate occurred at the same elution times on successive GC runs (Fig. 2b, c, e, f). The size of eluted peaks and the magnitude of neural responses were uncorrelated (Fig. 3c), and minor peaks elicited responses as robustly as large peaks (Fig. 2g, h).

Most urine-responsive neurons (80%, 83/104) were activated at a single time point, corresponding to a single eluted volatile urine component (Fig. 3a, b); 11% (11/104) responded at two time points and the remainder (10/104) responded at 3–5 time points (Supplementary Fig. 4). Averaging multiple runs to the same urine sample increased the signal-to-noise ratio of the responses, but did not reveal additional weak responses (Supplementary Fig. 5). The specific times of the responses were widely distributed over the course of the GC run. We identified at least 25 distinct response times, ranging from 2.5 min to 16 min after the start of the run (Fig. 3a). Several neurons responded to volatiles eluting at a specific time point, suggesting that we recorded from several mitral cells with similar or identical chemical selectivity across different animals. One particular subpopulation responded to a volatile eluting at ~508 s. As detailed below, these cells, observed in both males and females, responded to a component present exclusively in male urine (Fig. 3a, boxed region; cells that responded at this elution time but were not male-selective are not included in the box). These findings demonstrate that mitral cells are highly selective when confronted by a wide range of odorants at near-natural concentrations. This is not due to a weakened stimulus. Using a similar combination of SPME and GC-separation of urine volatiles, expert human olfactory analysts (Microanalytics) detected over 30 distinct odours eluting from the column (Supplementary Fig. 6). Mice probably detect a considerably larger repertoire of components.

### Sex-selectivity of mitral cell responses

To assess the sex- and strain-selectivity of the urine-selective neurons in the main olfactory bulb, we calculated a sex-selectivity index (SexSI) and a strain-selectivity index (StrSI) for 98 of the 104 neurons analysed using GC-E (six neurons were excluded owing to incomplete data). Both indices range from -1 (activated exclusively by female or C57Bl/6 urine) to +1 (activated exclusively by male or BALB/c urine) (Fig. 4g, h). Most neurons responded to urine from both strains, but 14% (14/98) responded much more robustly to BALB/c urine (from both sexes) than to C57Bl/6 urine (StrSI > 0.5). Neurons recorded in female mice showed an overall preference for male urine, with a median SexSI of 0.46; this bias was not observed among urine-responsive cells in male mice (median SexSI = 0.03). We found a single female-selective neuron in one male mouse (SexSI < -0.5), but no female-selective neurons in female mice, even though we recorded from almost three times as many urine-responsive neurons in females compared with males.

Within the population showing male preference (SexSI > 0.5), a subset of neurons responded exclusively to male urine (SexSI = 1.0, *n* = 14) (Fig. 4a, b). Three of these neurons were also tested with mouse urine from castrated males, and each failed to respond. Interestingly, 10/14 cells in this group were activated at virtually identical elution times (normalized time, 508 ± 1.8 s). Almost all (9/10) were recorded at similar stereotaxic coordinates in several different animals. These neurons showed an unusual firing pattern



**Figure 1** Urine-responsive mitral cells in the mouse main olfactory bulb. **a–c**, Responses of three single units in the mitral cell layer to stimulation by urine volatiles (2 s, indicated by black bar). An extracellular recording trace is shown for each cell, with the animal's breathing rhythm shown directly underneath. Raster plots and post-stimulus time histograms show responses to multiple stimulus presentations. Most responsive cells were rapidly excited by urine volatiles (**a**, **b**) with responses synchronized to the breathing rhythm, but a few cells were inhibited by urine volatiles (**c**). In the raster plots, responses to different types of urine are indicated by different colours: green, C57Bl/6 male; yellow, C57Bl/6 female; blue, BALB/c male; red, BALB/c female.

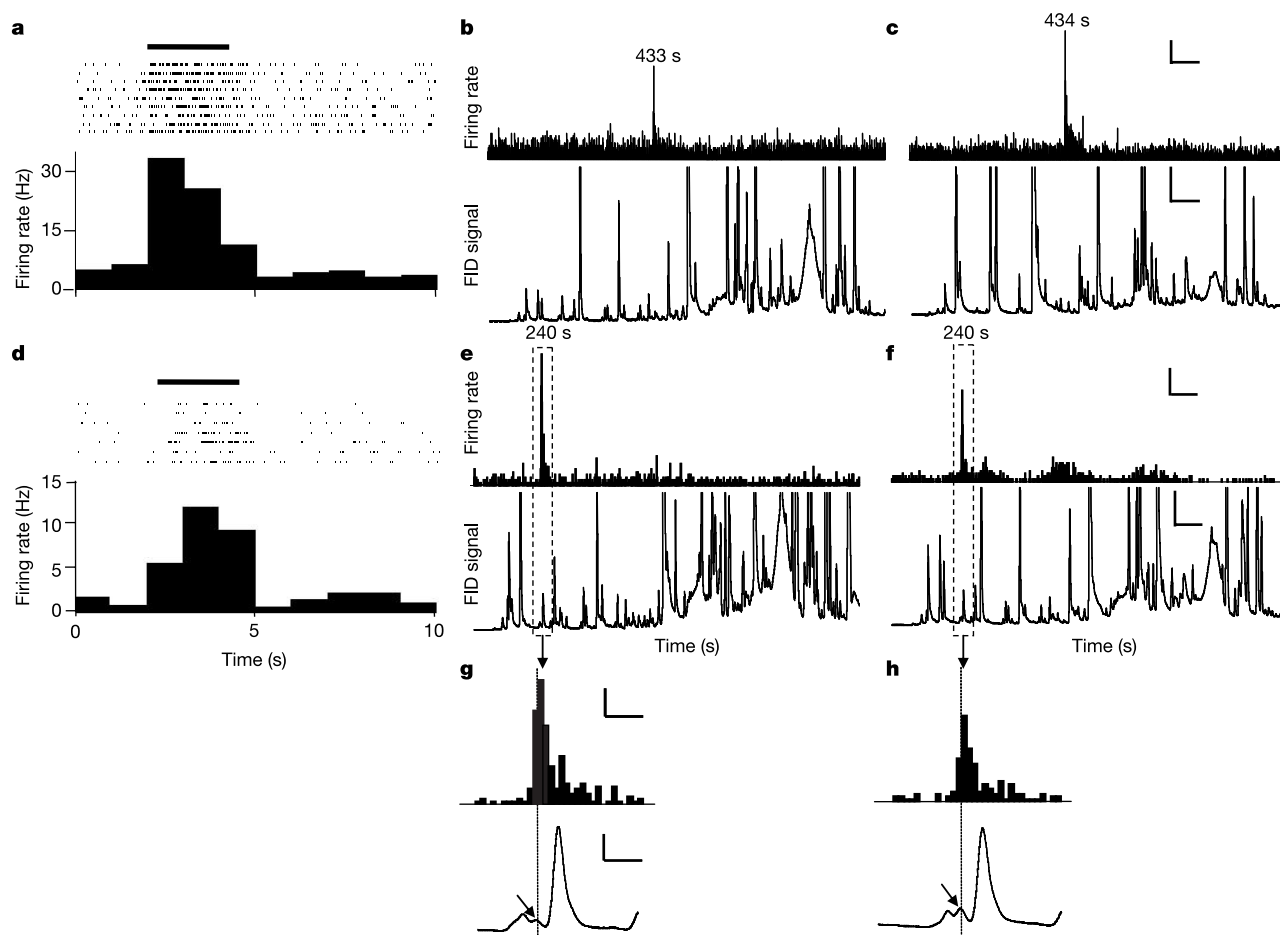


in response to the eluted stimulus, consisting of a sustained discharge that persisted well beyond the putative elution peak. The activation onset of these neurons (at 508 s) closely matched the elution time of a known mouse pheromone, 6-hydroxy-6-methyl-3-heptanone (HMH, ref. 21), which we identified by mass spectroscopy as a large peak in male urine eluting at  $500 \pm 6.8$  s. However, closer inspection of the GC profile revealed a small peak (partially embedded in the falling phase of the HMH peak) that reached maximum amplitude at  $\sim 508$  s. This peak was present in male urine GC profiles but absent in female and castrated male urine profiles (Fig. 4c–f). To establish that these neurons were responding to this second component, and not to HMH or a similar volatile, we sequentially tested responses to volatiles eluted on two different GC columns ( $n = 5$ ). When neurons were presented with volatiles eluted from our standard column (BP-5), which separates molecules according to boiling point, they responded at the expected 508 s. When we switched to a column that separates volatiles according to polarity (BP-20), the responses of the same neurons invariably shifted to 705 s, coincident with a small GC peak. This peak was completely distinct from the HMH peak, which, on this column, eluted at 375 s (Supplementary Fig. 7). The similarity of the neuronal responses to the elution of volatiles using two

columns that substantially change the presentation order of the stimuli also demonstrates that the responses of these neurons were independent of antecedent odours.

### A novel sex-specific compound in male urine

Data from high-resolution mass spectrometry indicated that the active compound had an exact mass corresponding to the formula  $C_2H_6S_2$  (molecular mass 93.9911). Analysis of sulphur isotope ratios also indicated that the compound contained two sulphur atoms. This peak showed fragment ions at  $m/q$  ratios of 61 (loss of SH), 47 ( $CH_3S$  or  $CH_2SH$ ) and 45 (base peak; HCS). Taken together, the data suggest three candidate structures: dimethyl disulphide, 1,2-ethanedithiol and (methylthio)methanethiol (MTMT). As the retention time of this unknown compound on both GC columns was clearly different from that of dimethyl disulphide or 1,2-ethanedithiol, MTMT was synthesized<sup>22</sup> (see Supplementary Methods). The retention times on both the BP-5 and BP-20 columns, and the mass spectral fragmentation patterns for synthetic MTMT precisely matched those for the unknown compound with  $m/q$  of 94 (Fig. 5c). Some structurally similar sulphur-containing compounds were also found in urine, including dimethyl disulphide ( $CH_3SSCH_3$ ), bis(methylthio)methane( $CH_3SCH_2SCH_3$ ) and



**Figure 2** Urine-responsive cells are activated by single components. **a**, A neuron's response to complete mouse urine volatiles. **b, c**, Firing rate (Hz) of the cell in **a**, as individual urine components are separated and presented using GC-E. Lower panels show gas chromatograms of C57Bl/6 male (**b**) and C57Bl/6 female (**c**) mouse urine. Larger peaks are truncated to enhance the visibility of smaller peaks. The cell responded at virtually the identical, single time point (433 and 434 s), in two independent GC-E recordings. Upper panel: vertical bar, 10 Hz; horizontal bar, 1 min. Lower panel: vertical

bar, 100 mV; horizontal bar, 1 min. **d–f**, Another urine-responsive mitral cell activated by a specific component of BALB/c male (**e**) or C57Bl/6 male (**f**) urine eluting at a different time (240 s) from that seen in **b, c**. Scale bars as above. **g, h**, Expanded representations of the regions indicated by dashed lines in **e** and **f**. The maximal cell response coincides with a small peak in the GC profiles (black arrows). Upper panel: vertical bar, 10 Hz; horizontal bar, 10 s. Lower panel: vertical bar, 20 mV; horizontal bar, 10 s.

methyl (methylthio)methyl disulphide ( $\text{CH}_3\text{SSCH}_2\text{SCH}_3$ ), but none of these elicited neuronal responses during GC runs. To humans, both synthetic MTMT and the native compound eluting at 508 s have a strong garlic or savory odour.

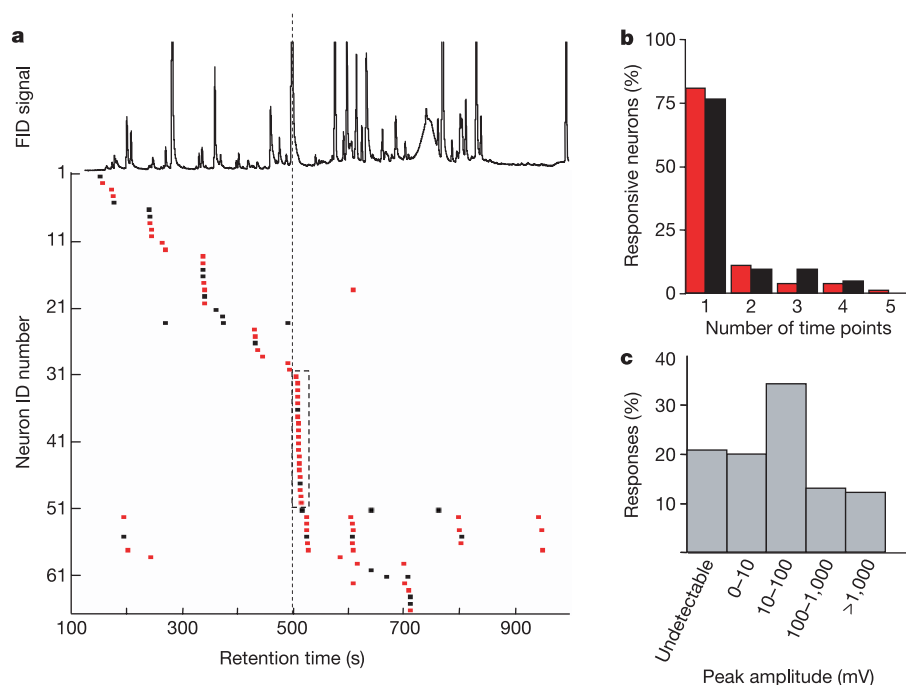
We next tested four neurons (three in females and one in males) that responded to eluted urine volatiles at 508 s, this time using synthetic MTMT. When presented either as a volatile or by elution from the GC, the cellular responses to MTMT precisely matched those of the endogenous compound (Fig. 5a, b). Using synthetic MTMT to construct a dose–response curve, we established the response threshold as 10 p.p.b. (parts per  $10^9$ , v/v) in water (Fig. 5d); male and female neurons had similar thresholds. By adding various amounts of MTMT to mouse urine, then performing SPME, we estimated that this molecule is naturally present in male mouse urine at a concentration of approximately 20 p.p.b. (v/v), but is almost undetectable in female urine. This value is similar to the behavioural threshold value reported for responses to another known volatile semiochemical that acts via the main olfactory system<sup>2</sup>. Native MTMT was detected in bladder urine, indicating that its presence in urine is not a consequence of bacterial metabolism, fecal contamination or compounds added to urine via glandular secretions. Thus we conclude that the male-specific responses in the main olfactory bulb primarily reflect activation of neurons by a previously unreported sulphur-containing compound secreted in male mouse urine.

### Behavioural responses to MTMT

Other sulphur-containing compounds, such as dimethyl disulphide and carbon disulphide, play important roles in rodent social behaviours<sup>23,24</sup>. To examine whether the presence of MTMT correlated with a behavioural role in sex discrimination, we took advantage of the fact that female mice strongly prefer to smell male (as opposed to female) urine, and that they largely ignore the urine from castrated males<sup>25</sup>, which has greatly reduced

concentrations of many volatiles including known pheromones and MTMT (we identified 112 peaks in the GC profile of intact C57Bl/6 male mouse urine but only 57 peaks in the urine of castrated C57Bl/6 male mice). Using a ‘Y’ maze assay, we found that all sexually experienced female mice (BALB/c,  $n = 14$ ) clearly preferred smelling urine from intact male mice to that from castrated male mice ( $P < 0.01$ , Fig. 6); most females (86%) also visited the arm of the maze containing the intact male urine first. To see whether MTMT could restore the attractiveness of urine from castrated males, we next presented female mice with a choice between ‘castrated’ urine and castrated urine to which 20 p.p.b. (v/v) MTMT was added. All but two females ( $n = 16$  out of 18) spent significantly more time investigating the castrated urine containing MTMT ( $P < 0.01$ , Fig. 6 and Supplementary Fig. 8). Female mice also showed a strong initial preference for the urine sample containing MTMT: 78% chose to sniff first at the arm with MTMT-containing urine. MTMT alone was ineffective in this behavioural paradigm, although female mice showed a slight preference for water containing 50 p.p.b. MTMT versus pure water. This implies the behavioural effects of MTMT are elicited in the context of other volatiles present in urine. The preference for castrated urine with added MTMT was not simply due to the presence of a novel odour. Acetophenone (another compound present in urine) added to castrated urine to a final concentration of 10 p.p.m. (at least tenfold higher than its endogenous concentration in urine) had no effect on female preference (Fig. 6). Although replenishing a single component—MTMT—does not fully restore female mouse preference for intact male mouse urine, it substantially increases the behavioural attractiveness of an otherwise uninteresting stimulus.

Given that female mice respond differently to MTMT depending on whether it is presented in water or castrated urine, we tested whether the presence of castrated urine volatiles modulated the responses of MTMT-responsive mitral cells (Supplementary



**Figure 3** Activation of urine-responsive neurons by single components. **a**, GC retention times at which individual neurons were activated. Normalized time points (squares) at which each of 66 urine-responsive cells were activated are aligned with a reference GC profile of male mouse urine. Red squares indicate female mice, black squares indicate male mice. Box indicates putative MTMT-responsive cells (see text and Fig. 5). **b**, Most

mitral cells recorded from both sexes respond at single elution times; a small fraction respond at multiple times (colour conventions as in **a**). **c**, The amplitude of the eluted peaks in the GC record are unrelated to their ability to activate urine-responsive neurons (386 responses, compared with the GC signal, in mV, from the FID detector)

Fig. 9). After isolating male-selective cells responsive to MTMT, we quantified the neuronal responses to 20 p.p.b. MTMT diluted in water, compared with 20 p.p.b. MTMT diluted in castrated urine ( $n = 3$ ). In all cases the responses were remarkably similar. Thus, the contextual dependency of female interest in MTMT is unlikely to arise from modulating the responses of MTMT-responsive cells themselves, but is probably a result of integration somewhere outside the olfactory bulb.

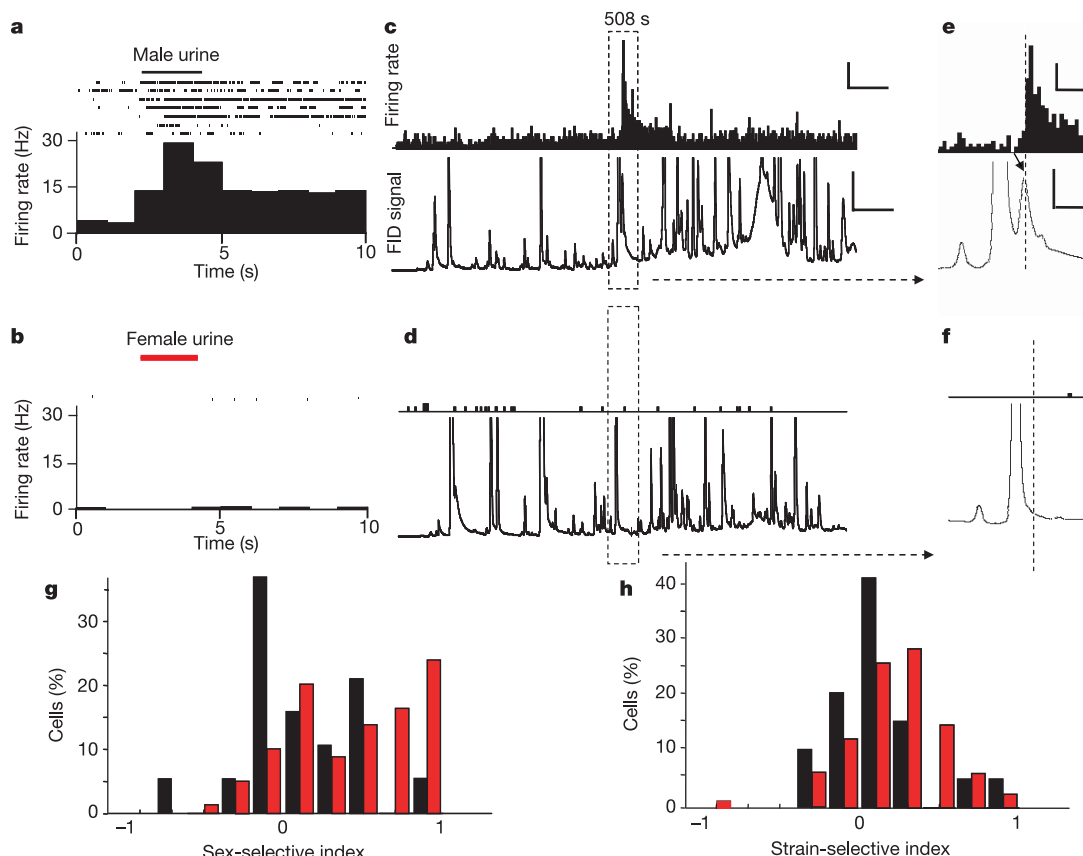
## Discussion

Volatiles from body secretions signal the presence of potential rivals and mates and motivate investigational behaviour in many mammalian species<sup>4</sup>. Even humans can reportedly distinguish sex on the basis of olfactory cues<sup>26,27</sup>. However, few of these sexual attractants are known. Here we show that a male-specific compound, MTMT, isolated by its ability to activate single neurons, considerably enhances the attractiveness of urine to female mice. Unlike previously described pheromones, MTMT is highly volatile, with a calculated vapour pressure at room temperature of ~15 torr. This is about 20-fold higher than exo-brevicomin and 2,000-fold higher than farnesene, two well-described pheromones that act via the accessory olfactory system<sup>28–30</sup>. MTMT seems likely to advertise the presence of a male from a distance, either as a signal to other males or to attract females. Levels of MTMT could serve as an indicator for the freshness of urine, especially since other sulphur-containing compounds in mouse urine, methanethiol ( $\text{CH}_3\text{SH}$ ) and dimethyl disulphide ( $\text{CH}_3\text{SSCH}_3$ ), can react with MTMT

to form methyl(methylthio)methyl disulphide and thereby gradually decrease MTMT concentration over time.

The finding that almost a third of urine-responsive neurons in both the male and female main olfactory bulb were activated by MTMT but that only a few cells were activated by any other single urine component (Fig. 3a) suggests either that multiple glomeruli are activated by MTMT, or that the glomerulus activated by MTMT is innervated by an unusually large number of mitral cells. These neurons could be associated with the 15–20 sexually dimorphic, 'atypical glomeruli' that have been implicated in the detection of biologically significant odours<sup>31–33</sup>. In other sensory systems, central representations of stimuli of particular significance have enlarged central representations, and the central representation of MTMT may represent an example of such an enhanced representation in the mammalian olfactory system, perhaps analogous to the enlarged representation of sex pheromones in the macroglomerular complex in insects<sup>34</sup>.

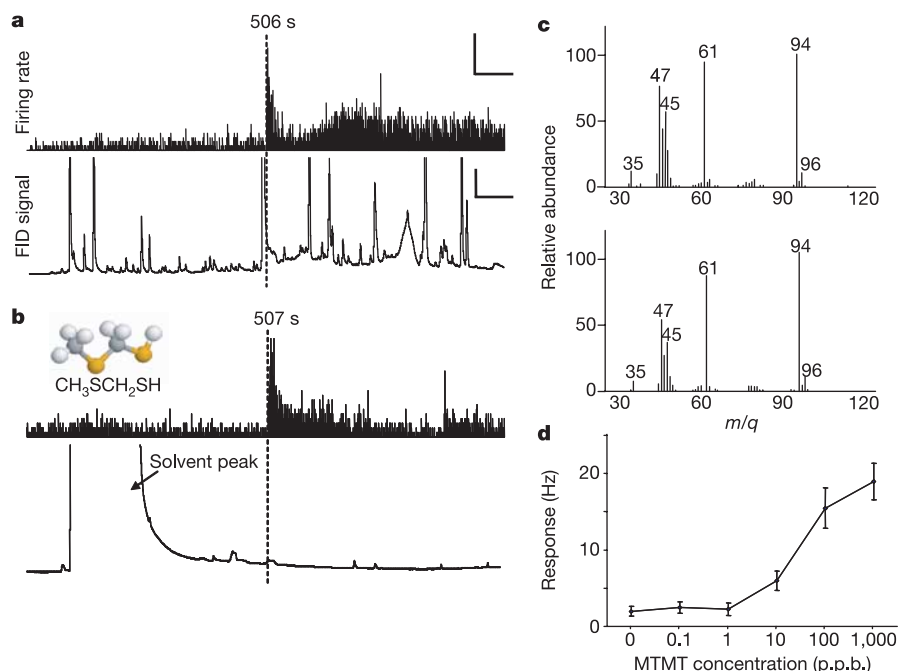
A persistent issue in the field of olfactory coding has been the odorant specificity of mitral cell activity<sup>35–39</sup>. The idea of 'specialist' and 'generalist' responsive neurons<sup>34</sup> has been invoked to explain the variability in molecular response ranges. However, the response range of mitral cells in mammals has historically been defined using relatively high odorant concentrations and restricted stimulus sets<sup>35,37</sup>. This does not necessarily relate to the discriminatory ability of mitral cells in biologically relevant contexts, where they are unlikely to encounter stimuli containing high, nearly identical concentrations of nearly identical stimuli. A more relevant question



**Figure 4** Sex- and strain-specific responses to urine volatiles. **a–f**, A mitral cell activated specifically by male urine volatiles (**a**) and not by female urine (**b**). Bar indicates time of odour delivery. **c, d**, Individual components of male BALB/c urine (**c**) elicit a response at 508 s (normalized response, conventions as in Fig. 3a), female C57BL/6 urine (**d**) elicits no response. Scale bars as in Fig. 2b–f. **e, f**, Enlarged portions of the GC profiles (boxed regions in **c** and **d**) show a small peak in male urine (**e**, black arrow) as the putative peak

activating the cell response; this peak is absent in female urine (**f**). Upper panel: vertical bar, 10 Hz; horizontal bar, 10 s. Lower panel: vertical bar, 100 mV; horizontal bar, 10 s. **g**, Sex-selectivity index shows a substantial subpopulation of neurons with an index score close to 1, indicating responses exclusively to male urine; such cells were less common in males (black bars). **h**, Most neurons responded to urine from both strains tested (colour conventions as previously).





**Figure 5** Male-specific neurons respond to a novel semiochemical. **a**, Activation of a mitral cell by a single component of BALB/c male mouse urine eluting at 506 s (normalized response). **b**, Synthetic MTMT (structure shown at upper left) elutes at the same time (507 s) and evokes an identical response (scales as in Fig. 2b–f). The small apparent size of the MTMT peak results from the low concentration used. **c**, Mass spectra of

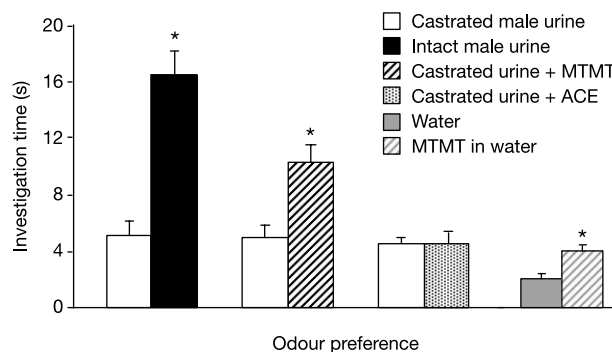
endogenous MTMT from mouse urine (upper panel) and synthetic MTMT (lower panel) obtained at the retention time corresponding to activation of the neuron in **a** and **b**. **d**, Dose–response curve to synthetic MTMT (averaged from four responsive cells, bars indicate mean  $\pm$  s.e.m.). Response threshold is approximately 10 p.p.b.

is the selectivity of mitral cells when confronted with arbitrarily large sets of distinct chemicals. In the case of projection neurons in the *Drosophila* antennal lobe (the fly analogue of mitral cells), imaging data suggest that most individual projection neurons respond robustly to only a small subset of stimuli<sup>40</sup>, but electrophysiological recordings of the same neurons indicate that most respond to a considerably broader range<sup>41</sup>. Broad tuning of mitral cells has also been reported in zebrafish<sup>36,39</sup>. Mice, however, have a far larger repertoire of olfactory receptors than either flies or fish<sup>42–44</sup>, and we see little evidence for such broad tuning; out of perhaps hundreds of possible volatile stimuli in urine, mouse mitral cells generally respond to just a single one.

Although this apparent selectivity could reflect a limitation of the SPME approach (such as a systematic failure to adsorb certain urine volatiles) we consider this unlikely. We found no neurons that responded to presentations of complete urine volatiles but subsequently failed to respond to at least one component during the GC run, as would be expected if we failed to capture the relevant molecular species by SPME. In addition, urine volatile profiles isolated by a theoretically unbiased technique (cryogenic trapping) identified a total of 98 chemicals<sup>45</sup>. Using SPME, we detected between 100 (in BALB/c female urine) and 131 volatile compounds (in BALB/c male urine) and identified several components (in addition to MTMT) that were not observed by cryogenic trapping. Moreover, the chemical families isolated by the two approaches were the same. Other studies comparing cryogenic trapping with SPME have concluded that the sensitivity, linearity and reliability of SPME are comparable to or superior than cryogenic trapping<sup>46–48</sup>. Although it remains possible that when only a single activity peak is observed by GC-E, a small number of additional compounds might also induce responses, mitral cells clearly respond to only a very small fraction of the volatile stimuli present in urine. This specificity could reflect a case of ‘specialist neurons’, but preliminary results suggest the same degree of specificity in response to components in a range of other natural stimuli as well as to very large libraries of

pure odorants (I. Davison, D. Y. L. and L. C. K., unpublished observations).

Although hundreds of compounds are present in natural stimuli, a much more restricted set of compounds may be responsible for constructing the olfactory percept. The olfactory image of a particular urine might be formed by the immediate, coordinate activation of a specific cohort of narrowly-tuned feature detectors<sup>49</sup>, rather than a time-dependent, emergent property of more moderately selective detectors<sup>50</sup>. Understanding how such a complex signal is represented provides insight not only into neural



**Figure 6** Synthetic MTMT enhances the attractiveness of urine to female mice. In a Y-maze preference test, females strongly prefer urine from intact males compared with castrated male mice, as indicated by investigation time of the different samples (shown as mean  $\pm$  s.e.m.). Addition of 20 p.p.b. synthetic MTMT to castrated urine roughly doubles investigation time. Much higher concentrations of acetophenone (ACE, 1 p.p.m. or 10 p.p.m.) have no effect. Although 50 p.p.b. MTMT in water slightly increased investigation time compared with water alone, the absolute investigation time is much lower than that of MTMT presented in a castrated urine background. Asterisks indicate statistically significant differences (Wilcoxon signed-ranks test,  $P < 0.01$ ).

strategies for the coding of social signals, but also suggests a paradigm for understanding more generally how natural olfactory scenes are represented in the olfactory bulb and beyond. □

## Methods

All animal experiments were performed according to a protocol approved by the Duke University Institutional Animal Care and Usage Committee.

## Electrophysiology

After initial anaesthesia (10:1 ketamine:xylazine, intraperitoneal injection), sexually mature (3–9-month-old) BALB/c or C57Bl/6 mice (Charles River Laboratory) were maintained on isoflurane (1–3% in 100% O<sub>2</sub>). The intersection of the sagittal suture and the sinus separating the olfactory bulbs and cortex was used as a reproducible zero point for mapping. Isolated single units were recorded with tungsten microelectrodes (1.5–3 MΩ, Microprobe Inc.), digitized (20 kHz) and analysed using Spike2 software (Cambridge Electronic Design Limited). Small electrolytic lesions (4 trains of 7 μA for 5 s, at 5 s intervals) were occasionally made and verified histologically.

## Data analysis

A cell was classified as urine-responsive if its firing rate during urine presentation changed by more than 30% from its spontaneous firing rate in the 2-s pre-stimulus interval. For each urine stimulus, responses to 2–5 stimulus presentations were averaged. We calculated a SexSI and StrSI as follows: SexSI = (male urine response) – (female urine response)/(male urine response + female urine response) and StrSI = (BALB/c urine response – C57Bl/6 urine response)/(BALB/c urine response + C57Bl/6 urine response).

Male and female urine responses are the sum of cell responses to BALB/c and C57Bl/6 animals of the respective sexes, and strain-selective responses are the sum of responses of both sexes to urines from different strains.

## Odour stimuli

Using custom-built metabolic cages, urine was collected from 5–10 animals into vials immersed in dry ice, and stored at –80 °C. Oestrous cycle was not controlled. Urine vapours (10% urine in dH<sub>2</sub>O) were presented in charcoal filtered room air (1.2 l min<sup>–1</sup>) using a TTL-triggered, 6-channel olfactometer. Later experiments used a robotic 64-channel olfactometer in which 100% urine vapour (0.2 l min<sup>–1</sup>) was diluted into a 2 l min<sup>–1</sup> flow stream. To concentrate volatiles from urine, a SPME fibre (2 cm–50/30 μm DVD/Carboxen/PDMS StableFlex, Supelco) was inserted for 3–6 h into the headspace of a 1.5-ml vial with a Teflon septum (Agilent) containing 150 μl of mouse urine (saturated with NaCl, mildly heated to 37 °C–40 °C and constantly stirred). Extracted volatiles were desorbed at the GC injection port (Agilent model 6980, 260 °C, 3 min) and separated on either a nonpolar BP-5 or a polar BP-20 column (SGE, 30 m, inside diameter 0.25 mm) using helium (1.2 ml min<sup>–1</sup>) as the carrier gas. Oven temperature was maintained at 50 °C for 2 min, ramped at 10 °C min<sup>–1</sup> to 200 °C, and at a further 30 °C min<sup>–1</sup> to a final temperature of 260 °C, maintained for 1 min. The column effluent was split equally and each half was routed to either a flame-ionization detector (FID, operating at 280 °C) or a mass spectrometry detector (Agilent model 5973, electron impact mode, 70 eV electron energy, 34–400 AMU at 3.93 scans s<sup>–1</sup>) or the animal's nose via a heated transfer line.

To compensate for slight shifts in retention times, the retention times for an unknown compound (UC) from different GC runs were adjusted according to the retention time (RT) of HMH as follows: adjusted RT of UC = (averaged RT of HMH/actual RT of HMH) × actual RT of UC. During GC-E experiments, cells were analysed only if their firing rate changed significantly (99% confidence level) from their mean firing rate, and their activity lasted longer than one breathing cycle. If multiple GC-E runs were performed, a given response had to be repeated at least twice.

## Behavioural analysis

Sexually experienced, light-cycle adapted (light on from 1800 to 0600 h) BALB/c female mice (*n* = 37) were used (during their subjective night, between 1200 and 1800 h) to examine the behaviour effects of synthetic MTMT in a urine preference test. Preference tests were conducted in a custom-made Y-maze (32 × 12 × 30 cm) with a sliding door regulating access to each arm. The test urine (50 μl, applied to a 1-cm<sup>2</sup> piece of filter paper) was placed at the bottom of a clean 1.5-ml eppendorf tube, which was in turn fitted into a circular port (1.3 cm inside diameter) located at the end of each arm. The mouse could sniff at the opening of the tube but was unable to contact its contents. Tests were performed in darkness (videotaped under infrared illumination). After habituation to the test apparatus (5 min), the animal was restricted to one arm, and two tubes containing filter paper with water only were inserted into the sniffing access ports. The animal was released and allowed to freely investigate for 5 min. The procedure was then repeated with urine samples placed on the filter paper (2-min trials). Videotapes were scored for the time spent sniffing each urine stimulus (snout oriented towards the opening and held within 1 cm of it) by an individual blind to the conditions being tested.

Received 23 September 2004; accepted 2 February 2005; doi:10.1038/nature03414.

Published online 20 February 2005.

1. Baldwin, B. A. & Shillito, E. E. The effects of ablation of the olfactory bulbs on parturition and maternal behaviour in Soay sheep. *Anim. Behav.* **22**, 220–223 (1974).
2. Schaal, B. *et al.* Chemical and behavioural characterization of the rabbit mammary pheromone. *Nature* **424**, 68–72 (2003).
3. Yamaguchi, M. *et al.* Distinctive urinary odors governed by the major histocompatibility locus of the mouse. *Proc. Natl Acad. Sci. USA* **78**, 5817–5820 (1981).
4. Brennan, P. A. & Keverne, E. B. Something in the air? New insights into mammalian pheromones.

*Curr. Biol.* **14**, R81–R89 (2004).

5. Dulac, C. & Torello, A. T. Molecular detection of pheromone signals in mammals: from genes to behaviour. *Nature Rev. Neurosci.* **4**, 551–562 (2003).
6. Andreolini, F., Jemiolo, B. & Novotny, M. V. Dynamics of excretion of urinary chemosignals in the house mouse (*Mus musculus*) during the natural estrous cycle. *Experientia* **43**, 998–1002 (1987).
7. Schwende, F. J., Wiesler, D., Jorgenson, J. W., Carmack, M. & Novotny, M. Urinary volatile constituents of the house mouse, *Mus musculus*, and their endocrine dependency. *J. Chem. Ecol.* **12**, 277–295 (1986).
8. Harvey, S., Jemiolo, B. & Novotny, M. Pattern of volatile compounds in dominant and subordinate male mouse urine. *J. Chem. Ecol.* **14**, 2061–2072 (1989).
9. Jemiolo, B., Xie, T. M., Andreolini, F., Baker, A. E. M. & Novotny, M. The t complex of the mouse: chemical characterization by urinary volatile profiles. *J. Chem. Ecol.* **17**, 353–367 (1990).
10. Schaefer, M. L., Young, D. A. & Restrepo, D. Olfactory fingerprints for major histocompatibility complex-determined body odors. *J. Neurosci.* **21**, 2481–2487 (2001).
11. Schaefer, M. L., Yamazaki, K., Osada, K., Restrepo, D. & Beauchamp, G. K. Olfactory fingerprints for major histocompatibility complex-determined body odors II: relationship among odor maps, genetics, odor composition, and behavior. *J. Neurosci.* **22**, 9513–9521 (2002).
12. Lodovichi, C., Belluscio, L. & Katz, L. C. Functional topography of connections linking mirror-symmetric maps in the mouse olfactory bulb. *Neuron* **38**, 265–276 (2003).
13. Belluscio, L., Lodovichi, C., Feinstein, P., Mombaerts, P. & Katz, L. C. Odorant receptors instruct functional circuitry in the mouse olfactory bulb. *Nature* **419**, 296–300 (2002).
14. Wadhams, L. Coupled gas chromatography-single cell recording: a new technique for use in the analysis of insect pheromones. *Z. Naturforsch.* **37c**, 947–952 (1982).
15. Stensmyr, M. C., Giordano, E., Balloi, A., Angioy, A. M. & Hansson, B. S. Novel natural ligands for *Drosophila* olfactory receptor neurones. *J. Exp. Biol.* **206**, 715–724 (2003).
16. Kayali-Sayadi, M. N., Bautista, J. M., Polo-Diez, L. M. & Salazar, I. Identification of pheromones in mouse urine by head-space solid phase microextraction followed by gas chromatography-mass spectrometry. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **796**, 55–62 (2003).
17. Louch, D., Motlagh, S. & Pawliszyn, J. Dynamics of organic compound extraction from water using liquid-coated fused silica fibers. *Anal. Chem.* **64**, 1187–1199 (1992).
18. Baek, H. H. & Kim, H. J. Solid phase microextraction-gas chromatography-olfactometry of soy sauce based on sample dilution analysis. *Food Sci. Biotech.* **13**, 90–95 (2004).
19. Lee, S. N., Kim, N. S. & Lee, D. S. Comparative study of extraction techniques for determination of garlic flavor components by gas chromatography-mass spectrometry. *Anal. Bioanal. Chem.* **337**, 749–756 (2003).
20. Torrens, J., Rui-Aumatell, M., Lopez-Tamames, E. & Buxaderas, S. Volatile compounds of red and white wines by headspace ETH solid-phase microextraction using different fibers. *J. Chromatogr. Sci.* **42**, 310–316 (2004).
21. Novotny, M. V. *et al.* A unique urinary constituent, 6-hydroxy-6-methyl-3-heptanone, is a pheromone that accelerates puberty in female mice. *Chem. Biol.* **6**, 377–383 (1999).
22. Schutte, L. One-step synthesis of dithiohemiacetals, a new class of compounds. *Tetrahedr. Lett.* **12**, 2321–2322 (1971).
23. Singer, A. G. *et al.* Dimethyl disulfide: an attractant pheromone in hamster vaginal secretion. *Science* **191**, 948–950 (1976).
24. Galef, B. G. Jr, Mason, J. R., Preti, G. & Bean, N. J. Carbon disulfide: a semiochemical mediating socially-induced diet choice in rats. *Physiol. Behav.* **42**, 119–124 (1988).
25. Scott, J. W. & Pfaff, D. W. Behavioral and electrophysiological responses of female mice to male urine odors. *Physiol. Behav.* **5**, 407–411 (1970).
26. Doty, R. L., Green, P. A., Ram, C. & Yankell, S. L. Communication of gender from human breath odors: relationship to perceived intensity and pleasantness. *Horm. Behav.* **16**, 13–22 (1982).
27. Wallace, P. Individual discrimination of humans by odor. *Physiol. Behav.* **19**, 577–579 (1977).
28. Novotny, M., Schwende, F. J., Wiesler, D., Jorgenson, J. W. & Carmack, M. Identification of a testosterone-dependent unique volatile constituent of male mouse urine: 7-exo-ethyl-5-methyl-6,8-dioxabicyclo[3.2.1]–3-octene. *Experientia* **40**, 217–219 (1984).
29. Leinders-Zufall, T. *et al.* Ultrasensitive pheromone detection by mammalian vomeronasal neurons. *Nature* **405**, 792–796 (2000).
30. Novotny, M., Harvey, S. & Jemiolo, B. Chemistry of male dominance in the house mouse, *Mus domesticus*. *Experientia* **46**, 109–113 (1990).
31. Shinoda, K., Shiotani, Y. & Osawa, Y. “Necklace olfactory glomeruli” form unique components of the rat primary olfactory system. *J. Comp. Neurol.* **284**, 362–373 (1989).
32. Werauga, E. *et al.* A sexually dimorphic group of atypical glomeruli in the mouse olfactory bulb. *Chem. Senses* **26**, 7–15 (2001).
33. Lin, W., Arellano, J., Slotnick, B. & Restrepo, D. Odors detected by mice deficient in cyclic nucleotide-gated channel subunit A2 stimulate the main olfactory system. *J. Neurosci.* **24**, 3703–3710 (2004).
34. Hildebrand, J. G. & Shepherd, G. M. Mechanisms of olfactory discrimination: converging evidence for common principles across phyla. *Annu. Rev. Neurosci.* **20**, 595–631 (1997).
35. Mori, K. Relation of chemical structure to specificity of response in olfactory glomeruli. *Curr. Opin. Neurobiol.* **5**, 467–474 (1995).
36. Friedrich, R. W. & Laurent, G. Dynamic optimization of odor representations by slow temporal patterning of mitral cell activity. *Science* **291**, 889–894 (2001).
37. Motokizawa, F. Odor representation and discrimination in mitral/tufted cells of the rat olfactory bulb. *Exp. Brain Res.* **112**, 24–34 (1996).
38. Tanabe, T., Iino, M. & Takagi, S. F. Discrimination of odors in olfactory bulb, pyriform-amygdaloid areas, and orbitofrontal cortex of the monkey. *J. Neurophysiol.* **38**, 1284–1296 (1975).
39. Friedrich, R. W., Habermann, C. J. & Laurent, G. Multiplexing using synchrony in the zebrafish olfactory bulb. *Nature Neurosci.* **7**, 862–871 (2004).
40. Wang, J. W., Wong, A. M., Flores, J., Vossahl, L. B. & Axel, R. Two-photon calcium imaging reveals an odor-evoked map of activity in the fly brain. *Cell* **112**, 271–282 (2003).
41. Wilson, R. I., Turner, G. C. & Laurent, G. Transformation of olfactory representations in the *Drosophila* antennal lobe. *Science* **303**, 366–370 (2004).
42. Vossahl, L. B., Amrein, H., Morozov, P. S., Rzhetsky, A. & Axel, R. A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell* **96**, 725–736 (1999).
43. Clyne, P. J. *et al.* A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. *Neuron* **22**, 327–338 (1999).
44. Ngai, J., Dowling, M. M., Buck, L., Axel, R. & Chess, A. The family of genes encoding odorant receptors in the channel catfish. *Cell* **72**, 657–666 (1993).

45. Miyashita, K. & Robinson, A. B. Identification of compounds in mouse urine vapor by gas chromatography and mass spectrometry. *Mech. Ageing Dev.* **13**, 177–184 (1980).
46. Bocchini, P., Andalo, C., Bonfiglioli, D. & Galletti, G. C. Solid-phase microextraction gas chromatography/mass spectrometric analysis of volatile organic compounds in water. *Rapid Commun. Mass Spectrom.* **13**, 2133–2139 (1999).
47. Contarini, G. & Povolito, M. Volatile fraction of milk: comparison between purge and trap and solid phase microextraction techniques. *J. Agric. Food Chem.* **50**, 7350–7355 (2002).
48. Fu, S. G., Yoon, Y. & Bazemore, R. Aroma-active components in fermented bamboo shoots. *J. Agric. Food Chem.* **50**, 549–554 (2002).
49. Brody, C. D. & Hopfield, J. J. Simple networks for spike-timing-based computation, with application to olfactory processing. *Neuron* **37**, 843–852 (2003).
50. Laurent, G. Olfactory network dynamics and the coding of multidimensional signals. *Nature Rev. Neurosci.* **3**, 884–895 (2002).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank R. Axel, M. Ehlers, R. Mooney, D. Fitzpatrick and members of the Katz laboratory for critical comments on the manuscript; J. Jin, R. Irving, G. Dubay and L. Nielsen for technical assistance; and X. Han for critical assistance with chemical analysis. This work is supported by the NIH (NICDS) (L.C.K.), the Broad Foundation (D.Y.L.), NSF (E.B.), the Petroleum Research Fund, administered by the American Chemical Society (E.B.), and the Berryman Institute (E.B.). L.C.K. is an Investigator in the Howard Hughes Medical Institute.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to D.Y.L. ([dayulin@neuro.duke.edu](mailto:dayulin@neuro.duke.edu)).



# Filamentary structure on the Sun from the magnetic Rayleigh–Taylor instability

Hiroaki Isobe<sup>1</sup>, Takehiro Miyagoshi<sup>1</sup>, Kazunari Shibata<sup>1</sup> & Takaaki Yokoyama<sup>2</sup>

<sup>1</sup>Kwasan and Hida Observatories, Kyoto University, Yamashina, Kyoto 607-8471, Japan

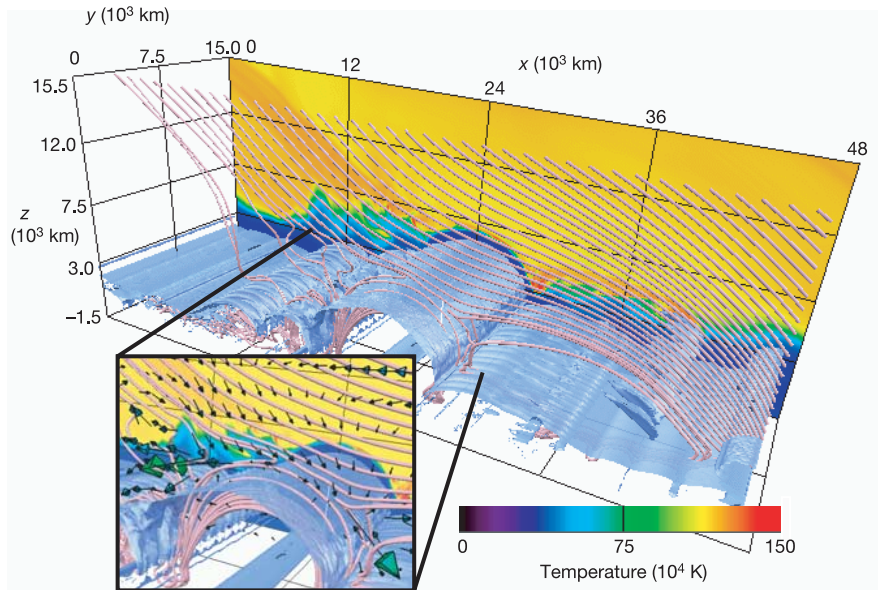
<sup>2</sup>Department of Earth and Planetary Science, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Magnetic flux emerges from the solar surface as dark filaments connecting small sunspots with opposite polarities<sup>1–3</sup>. The regions around the dark filaments are often bright in X-rays and are associated with jets<sup>4–6</sup>. This implies plasma heating and acceleration, which are important for coronal heating. Previous two-dimensional simulations of such regions showed that magnetic reconnection between the coronal magnetic field and the emerging flux produced X-ray jets and flares, but left unresolved the origin of filamentary structure and the intermittent nature of the heating. Here we report three-dimensional simulations of emerging flux showing that the filamentary structure arises spontaneously from the magnetic Rayleigh–Taylor instability<sup>7,8</sup>, contrary to the previous view that the dark filaments are isolated bundles of magnetic field that rise from the photosphere carrying the dense gas<sup>9–11</sup>. As a result of the magnetic Rayleigh–Taylor

instability, thin current sheets are formed in the emerging flux, and magnetic reconnection occurs between emerging flux and the pre-existing coronal field in a spatially intermittent way. This explains naturally the intermittent nature of coronal heating and the patchy brightenings in solar flares.

We performed three-dimensional magnetohydrodynamic simulations of solar emerging flux and its reconnection with pre-existing magnetic field in the corona, with the highest resolution ever achieved. The numerical code is basically the same as that used in our previous two-dimensional simulations<sup>10,12,13</sup>, but extended to include variations in the third ( $y$ ) direction. The simulation box is designed to model the upper convection zone (convectively unstable), photosphere/chromosphere ( $10^4$  K), and corona ( $10^6$  K). The box size is  $0 < x < 48,000$ ,  $0 < y < 15,000$ ,  $-1,500 < z < 19,500$  (units are kilometres throughout), which is resolved by  $800 \times 400 \times 620$  grid points;  $z$  is the vertical direction and  $z = 0$  corresponds to the level of the photosphere. The initial state is hydrostatic. To simulate an emerging flux region (EFR), a horizontal magnetic flux sheet is placed in the convection zone. A small perturbation is then imposed on the flux sheet within a finite domain ( $22,500 < x < 25,500$ ), to excite the Parker instability<sup>10,11</sup>. The magnetic field in the corona is oblique and antiparallel to the flux sheet. The boundary conditions are: a rigid wall at the bottom, a free boundary at the top, and periodic at the  $x$  and  $y$  boundaries. Following previous studies of fast magnetic reconnection, we adopt an anomalous resistivity model because localized resistivity is necessary for fast reconnection to occur<sup>14,15</sup>.

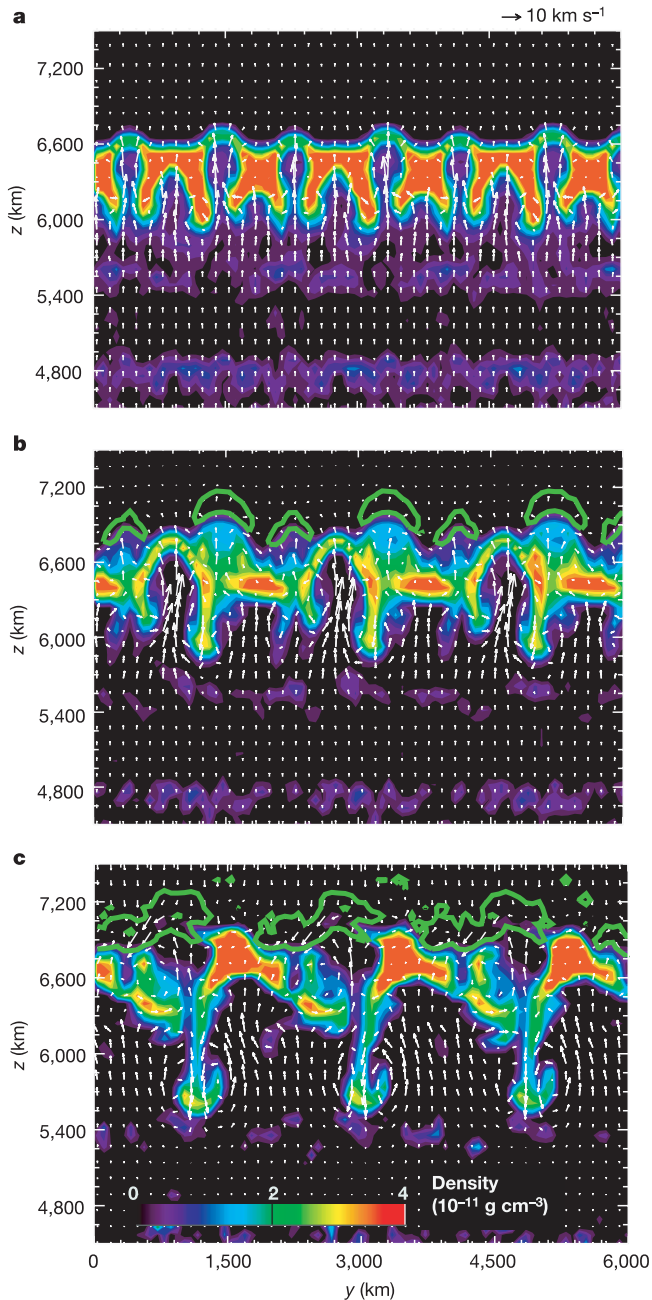
Figure 1 shows the three-dimensional visualization of the simulation result, showing the isosurface of magnetic field strength  $|B|$ , temperature distributions at the  $y$  boundary, a set of representative



**Figure 1** Three-dimensional visualization of the simulation result at  $t = 2,500$  s. Isosurfaces of the magnetic field strength  $|B| = 40$  G (blue transparent surfaces), temperature distribution on a boundary (colour), and a set of magnetic field lines near  $y = 2,400$  km (pink tubes) are shown. The simulation box contains the corona ( $10^6$  K), the chromosphere/photosphere ( $10^4$  K) and the convection zone (convectively unstable gas layer). The initial magnetic field of the flux sheet is given by;  $\mathbf{B}_{\text{sheet}}(z) = (B_s(z), 0, 0)$  ( $-1,200 < z < -600$ ).  $B_s(z)$  is chosen so that the plasma  $\beta$  (the ratio of gas pressure to magnetic pressure) in the initial flux sheet is 4 in the range  $-1,200 < z < -600$ ,  $B_s(-600) \approx 1,200$  G and  $B_s(-1,200) \approx 2,100$  G. We also impose a uniform oblique field that has a component antiparallel with the initial flux sheet:  $\mathbf{B}_{\text{corona}} = (-B_c \cos \theta, 0, B_c \sin \theta)$ , where  $B_c = 26$  G and  $\theta = 150^\circ$ . The plasma  $\beta$  in the initial corona is about 2.6. The initial perturbation is imposed on the  $z$  component of the

velocity in the range  $22,500 < x < 25,500$  and  $-1,200 < z < -600$ ,  $V_z = V_p \cos(2\pi(x - x_{\text{mid}})/\lambda)$ , where  $V_p = 0.5 \text{ km s}^{-1}$ ,  $x_{\text{mid}} = 24,000$  km and  $\lambda = 6,000$  km. When the emerging flux enters the corona, current sheets form near  $(x, z) = (25,000, 7,000)$  where the emerging flux and coronal magnetic field is most antiparallel. The current sheet is recognized as the thin high-temperature ( $1.5 \times 10^6$  K) region. At this time the plasma  $\beta$  near the reconnection region is about 0.001 in the emerging flux and about 1 in the corona. The inset shows a close-up of the vicinity of magnetic reconnection region. The arrows indicate the velocity field. A pair of V-shaped field lines and oppositely directed plasma flows are the signature of magnetic reconnection. Although the morphology and dynamics do not vary significantly along the  $y$  direction, three-dimensional structure is seen in the isosurface of  $|B|$ .

magnetic field lines, and the velocity field at  $t = 2,500$  s. The omega-shaped structure seen in the  $|\mathbf{B}|$  isosurfaces and field lines is the emerging flux. As the emerging flux pushes the coronal field, thin current concentrations with a high current density are formed near

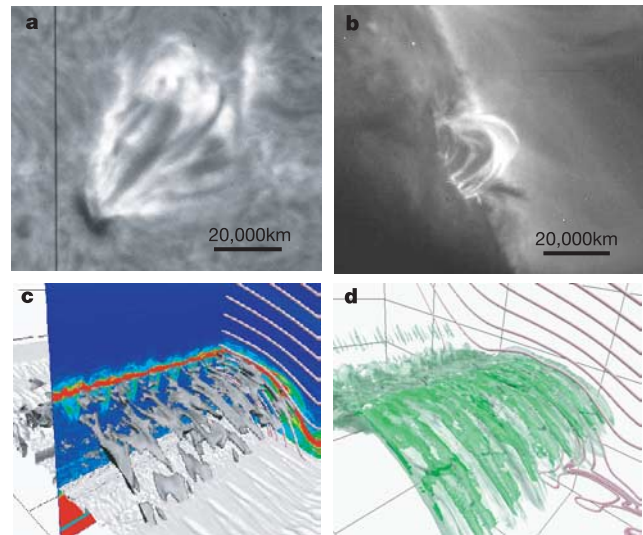


**Figure 2** Temporal evolution of the mass density (colour), velocity (arrows) and resistivity (green contour) on a  $y$ - $z$  plane near the reconnection points. The times are  $t = 1,976$  s (**a**),  $t = 2,028$  s (**b**) and  $t = 2,106$  s (**c**), and the  $y$ - $z$  plane is at  $x = 24,400$  km. Only the velocity components on the plane (that is, the  $y$  and  $z$  components) are shown. The contour level of resistivity is  $6 \times 10^{-11} \text{ cm}^2 \text{ s}^{-1}$ , showing the region where anomalous resistivity sets in. The functional form of the resistivity is taken to be  $\eta = \eta_0$  for  $v_d < v_c$  and  $\eta = \eta_0 + \eta_0 (v_d/v_c - 1)^2$  for  $v_d > v_c$ , where  $v_d = J/\rho$  is the non-dimensional electron drift velocity,  $\rho$  is the mass density,  $J$  is the current density, and  $v_c$  is the threshold above which the anomalous resistivity sets in. The inclusion of constant background resistivity  $\eta_0 = 3 \times 10^{-11} \text{ cm}^2 \text{ s}^{-1}$  is for numerical stability and does not significantly affect the dynamics of the emerging flux because the timescale of the slow reconnection/diffusion by the constant resistivity is much larger than the dynamical timescale of the emerging flux.

$(x, z) = (25,000, 7,000)$  and anomalous resistivity sets in there. Then fast reconnection of the magnetic field occurs, which is manifested by sets of highly bent field lines that, like stretched rubber bands, accelerate plasma in opposite directions (Fig. 1, inset).

The morphology and dynamics are similar in the  $y$  direction. More detail is given in previous two-dimensional work<sup>12,13</sup>. However, note that the three-dimensional structure is seen in the isosurfaces of  $|\mathbf{B}|$ . Although no variation along the  $y$  direction is imposed explicitly in the initial conditions, there is numerical noise of the order of  $10^{-7}$  or less, which grows nonlinearly and forms a three-dimensional structure. We found that the growth of the three-dimensional structure is due to the magnetic Rayleigh–Taylor instability<sup>7,8</sup>. Figure 2 shows the mass density, velocity and resistivity distributions on the  $y$ - $z$  plane at  $x = 24,400$  km. The denser part near  $z = 6,000$ – $7,000$  km corresponds to the top of the emerging flux. This top-heavy configuration is unstable for the Rayleigh–Taylor instability. As it grows through the nonlinear phase, the sinking spikes form mushroom-like structures due to the Kelvin–Helmholtz instability<sup>7,8</sup>. Bending of the magnetic field, which is nearly perpendicular to the  $y$ - $z$  plane, is stabilized by the magnetic tension. That is, only interchange modes grow. The resultant structure therefore consists of ‘filaments’ along the magnetic field.

A comparison with solar observations is given in Fig. 3. Figure 3a shows an H $\alpha$  image of an EFR. The arch filaments connect a sunspot (lower left) with a bright region called plage (upper right). The length and width of the filaments are 10,000–30,000 km and 1,000–3,000 km, respectively. **b**, Extreme ultraviolet image of an EFR near the solar limb taken by the TRACE satellite on 8 June 1998. Bright loops are hot ( $10^6$  K) and seen in emission, whereas dark loops are cold ( $10^4$  K) and seen in absorption. A dark jet is being ejected from the lower side of the loops. **c**, Current density distribution on an  $x$ - $z$  plane and a  $y$ - $z$  plane (colour) as well as isosurfaces of mass density  $\rho = 2 \times 10^{-11} \text{ g cm}^{-3}$  (grey surface) at  $t = 2,200$  s. The isosurfaces of  $\rho$  trace the upper chromosphere and the dense filaments in the corona;  $z > 4,000$  km is drawn in dark grey, to highlight the filaments in the corona. The red parts correspond to the strongest current region between the emerging flux and the coronal field. The green spikes below the red parts are the filamentary current sheets in the emerging flux itself, formed in the periphery of sinking spikes (mushrooms) as a result of the Rayleigh–Taylor instability. **d**, Isosurface of current density  $6 \times 10^{-2} \text{ A m}^{-2}$  (green) and  $6 \times 10^{-3} \text{ A m}^{-2}$  (transparent).



**Figure 3** Comparison of the simulation result and observations. **a**, H $\alpha$  image of an EFR taken with the Domeless Solar Telescope at Hida observatory, Kyoto University, on 4 October 1991 (courtesy of H. Kurokawa and A. Adamura). The dark filaments connect a sunspot (lower left) with a bright region called plage (upper right). The length and width of the filaments are 10,000–30,000 km and 1,000–3,000 km, respectively. **b**, Extreme ultraviolet image of an EFR near the solar limb taken by the TRACE satellite on 8 June 1998. Bright loops are hot ( $10^6$  K) and seen in emission, whereas dark loops are cold ( $10^4$  K) and seen in absorption. A dark jet is being ejected from the lower side of the loops. **c**, Current density distribution on an  $x$ - $z$  plane and a  $y$ - $z$  plane (colour) as well as isosurfaces of mass density  $\rho = 2 \times 10^{-11} \text{ g cm}^{-3}$  (grey surface) at  $t = 2,200$  s. The isosurfaces of  $\rho$  trace the upper chromosphere and the dense filaments in the corona;  $z > 4,000$  km is drawn in dark grey, to highlight the filaments in the corona. The red parts correspond to the strongest current region between the emerging flux and the coronal field. The green spikes below the red parts are the filamentary current sheets in the emerging flux itself, formed in the periphery of sinking spikes (mushrooms) as a result of the Rayleigh–Taylor instability. **d**, Isosurface of current density  $6 \times 10^{-2} \text{ A m}^{-2}$  (green) and  $6 \times 10^{-3} \text{ A m}^{-2}$  (transparent).



(EUV) image of another EFR near the solar limb. Bright loops are hot ( $\sim 10^6$  K) and dark loops are cold ( $\sim 10^4$  K). The coexistence of many hot and cold loops is evidence for spatially intermittent heating. Jets are ejected from the lower footpoints of the loops, indicating the occurrence of magnetic reconnection between the emerging flux and the pre-existing coronal field<sup>12,13,16</sup>.

The grey surfaces in Fig. 3c are isosurfaces of mass density in the simulation. The dark grey surfaces trace the cold and dense plasma in the corona. They appear as isolated filaments that are very similar to the observed dark filaments. The colours on the vertical planes are the current density distributions; they show many small current sheets created in the emerging flux by the Rayleigh–Taylor instability. Dissipation of the filamentary current sheets leads to heating of the plasma in the periphery of the dense filaments and formation of a system of hot and cold loops existing alternately. Thus, the intermittent heating of the emerging flux indicated by the EUV observations is explained naturally.

Observational evidence of the formation of current sheets in emerging flux has recently been reported<sup>17</sup>, supporting the idea that the corona is heated by dissipation of current sheets<sup>18</sup>. Shuffling of magnetic field by photospheric flows is often considered to be the driving mechanism of current formation in the corona<sup>18–20</sup>. Our numerical simulations indicate that the Rayleigh–Taylor instability can also form current sheets in EFRs, and can also explain why the heating occurs in an intermittent way.

Such structure formation by the Rayleigh–Taylor and Kelvin–Helmholtz instabilities occurs in diverse situations in natural phenomena such as supernova remnants<sup>21</sup> and technological applications such as laboratory laser implosion<sup>7</sup>. As for the origin of the arch filaments in the solar EFR, it is also possible that the emerging flux is already fragmented in the photosphere by, for example, turbulent convection. Even so, however, when the fragmented magnetic flux emerges into the upper atmosphere, magnetic energy becomes dominant and hence the fragmented magnetic flux tubes

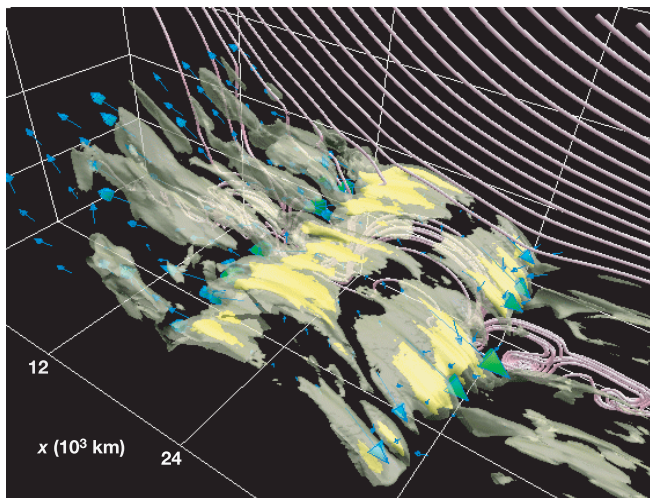
expand horizontally and fill the space between them, developing the almost two-dimensional structure except at the edge of an EFR<sup>11</sup>. Therefore, basically the same physical processes should occur in more complicated three-dimensional structures.

The formation of filamentary structure has significant consequences for the behaviour of magnetic reconnection. The green lines in Fig. 2 show resistivity contours. As the Rayleigh–Taylor instability grows through the nonlinear phase, the current density is enhanced in the rising part by upward forcing. Because the mass density is lower in the rising parts, the resistivity is locally enhanced in those parts owing to the anomalous resistivity model (see the legend to Fig. 2). Fast reconnection therefore starts only in the rising parts; the plasma is then accelerated and ejected from the current sheet by the reconnected magnetic field. The mass density and gas pressure in the diffusion region decrease, inducing plasma flows towards the diffusion region that pinch the current sheet. The current density and hence resistivity are thereby enhanced. It is a kind of nonlinear instability, known as spontaneous fast reconnection<sup>14,22</sup>. In the present simulation, fast reconnection is initiated locally by the Rayleigh–Taylor instability, and the fast reconnection enhances the growth of the instability by inducing plasma flows. Therefore the Rayleigh–Taylor instability and the fast magnetic reconnection are nonlinearly coupled, resulting in spatially intermittent, ‘patchy’ reconnection.

Figure 4 shows the isosurfaces of velocity  $|V|$  that trace the fast outflows from the reconnecting current sheets and also plasma flows moving upwards along the ambient field lines. Because of intermittent reconnection, many distinctive narrow outflows (jets) are identified. Although the present simulation cannot reproduce the electromagnetic radiation qualitatively, the patchy reconnection should produce patchy brightenings, which are actually observed in EFRs (see Fig. 3).

The intermittent, filamentary structures and patchy brightenings are often observed not only in EFRs but also in various solar phenomena<sup>23–25</sup>. Considering the ubiquity of the filamentary/patchy structures, it is likely that, if the reconnecting current sheet is not in mechanical equilibrium and the densities of both the sides are different, patchy reconnection can occur by the coupling of anomalous resistivity with the Rayleigh–Taylor instability, as demonstrated in our high-resolution simulation. We conjecture that the same mechanism can be applied to explain the origin of filamentary structure and patchy brightenings in various dynamic phenomena in astrophysical, space and laboratory plasmas.  $\square$

Received 3 September 2004; accepted 26 January 2005; doi:10.1038/nature03399.



**Figure 4** Three-dimensional visualization of the simulation result, showing the structure of plasma flows from reconnection points. The isosurfaces of velocity  $|V| = 120 \text{ km s}^{-1}$  (yellow) and  $60 \text{ km s}^{-1}$  (transparent), velocity vector (arrows), and a set of field lines are shown. The isosurface of  $|V| = 120 \text{ km s}^{-1}$  traces the fast outflows ejected from reconnecting current sheets, whereas the  $|V| = 60 \text{ km s}^{-1}$  surface traces the jets moving upwards along the ambient field lines. Because of the patchy distribution of the reconnection points, many narrow jets are ejected from the current sheet. This may correspond to the fine structures observed in coronal jets. Although the present simulation cannot quantitatively reproduce the electromagnetic radiation observed, impulsive heating by the patchy reconnection should produce patchy brightenings in the chromosphere, the transition region and the corona because energy transport from the reconnection site by thermal conduction and high energy particles are only along the magnetic field.

1. Bruzek, A. On arch-filament system in spotgroups. *Sol. Phys.* **2**, 451–461 (1967).
2. Zirin, H. Fine structure of solar magnetic fields. *Sol. Phys.* **22**, 34–48 (1972).
3. Frazier, E. N. The magnetic structure of arch filament system. *Sol. Phys.* **26**, 130–141 (1972).
4. Golub, L., Rosner, R., Vaiana, G. S. & Weiss, N. O. Solar magnetic fields: the generation of emerging flux. *Astrophys. J.* **243**, 309–316 (1981).
5. Yoshimura, K. & Kurokawa, H. Causal relations between H $\alpha$  loop emergences and soft X-ray brightenings. *Astrophys. J.* **517**, 964–976 (1999).
6. Shibata, K. *et al.* Observations of X-ray jets with the YOHKOH Soft X-ray Telescope. *Publ. Astron. Soc. Jpn.* **44**, L173–L179 (1992).
7. Sharp, D. H. An overview of Rayleigh–Taylor instability. *Physica* **12D**, 3–18 (1984).
8. Cattaneo, F. & Hughes, D. H. The nonlinear break up of a magnetic layer: instability to interchange modes. *J. Fluid Mech.* **196**, 323–344 (1988).
9. Zwaan, C. The emergence of magnetic flux. *Sol. Phys.* **100**, 397–414 (1985).
10. Shibata, K. *et al.* Nonlinear Parker instability of isolated magnetic flux in a plasma. *Astrophys. J.* **338**, 471–492 (1989).
11. Matsumoto, R., Tajima, T., Shibata, K. & Kaisig, M. Three-dimensional magnetohydrodynamics of the emerging magnetic flux in the solar atmosphere. *Astrophys. J.* **414**, 357–371 (1993).
12. Yokoyama, T. & Shibata, K. Magnetic reconnection as the origin of X-ray jets and H-alpha surges in the Sun. *Nature* **375**, 42–44 (1995).
13. Yokoyama, T. & Shibata, K. Numerical simulation of solar coronal X-ray jets based on the magnetic reconnection model. *Publ. Astron. Soc. Jpn.* **48**, 353–376 (1996).
14. Ugai, M. Computer studies on development of the fast reconnection mechanism for different resistivity models. *Phys. Fluids B* **4**, 2953–2963 (1992).
15. Yokoyama, T. & Shibata, K. What is the condition for fast magnetic reconnection? *Astrophys. J.* **436**, L197–L200 (1994).
16. Miyagoshi, T. & Yokoyama, T. Magnetohydrodynamic numerical simulations of solar X-ray jets based on the magnetic reconnection model that includes chromospheric evaporation. *Astrophys. J.* **593**, L133–L136 (2003).



17. Solanki, S. K. *et al.* Three-dimensional magnetic field topology in a region of solar coronal heating. *Nature* **425**, 692–695 (2003).
18. Parker, E. N. Nanoflare and the solar X-ray corona. *Astrophys. J.* **330**, 474–479 (1988).
19. Galsgaard, K. & Nordlund, A. Heating and activity of the solar corona. 1. Boundary shearing of an initially homogeneous magnetic field. *J. Geophys. Res.* **101**, 13445–13460 (1996).
20. Karpen, J. T., Antiochos, S. K. & DeVore, C. R. Reconnection-driven current filamentation in solar arcades. *Astrophys. J.* **460**, L73–L76 (1996).
21. Hachisu, I., Matsuda, T., Nomoto, K. & Shigeyama, T. Nonlinear growth of Rayleigh–Taylor instabilities and mixing in SN 1987A. *Astrophys. J.* **358**, L57–L61 (1990).
22. Ugai, M. & Shimizu, T. Computer studies on the spontaneous fast reconnection mechanism in three dimensions. *Phys. Plasmas* **3**, 853–862 (1996).
23. Kitahara, T. & Kurokawa, H. High-resolution observation and detailed photometry of a great H $\alpha$  two-ribbon flare. *Sol. Phys.* **125**, 321–332 (1990).
24. Innes, D. E., McKenzie, D. E. & Wang, T. SUMER spectral observations of post-flare supra-arcade inflows. *Sol. Phys.* **217**, 247–265 (2003).
25. Asai, A., Yokoyama, T., Shimojo, M. & Shibata, K. Downflow motions associated with impulsive nonthermal emissions observed in the 2002 July 23 solar flare. *Astrophys. J.* **605**, L77–L80 (2004).

**Acknowledgements** The authors thank N. O. Weiss, A. Asai and D. H. Brooks for comments. Use of TRACE data is acknowledged. This work was supported by the Japan–UK Cooperation Science Program of the JSPS (Principal investigators K.S. and N. O. Weiss) and a Grant-in-Aid for the 21st Century COE ‘Centre for Diversity and Universality in Physics’ from MEXT, Japan. The numerical computation was performed on the Earth Simulator.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to H.I. (isobe@kwasan.kyoto-u.ac.jp).

## Spatial quantum noise interferometry in expanding ultracold atom clouds

Simon Fölling, Fabrice Gerbier, Artur Widera, Olaf Mandel, Tatjana Gericke & Immanuel Bloch

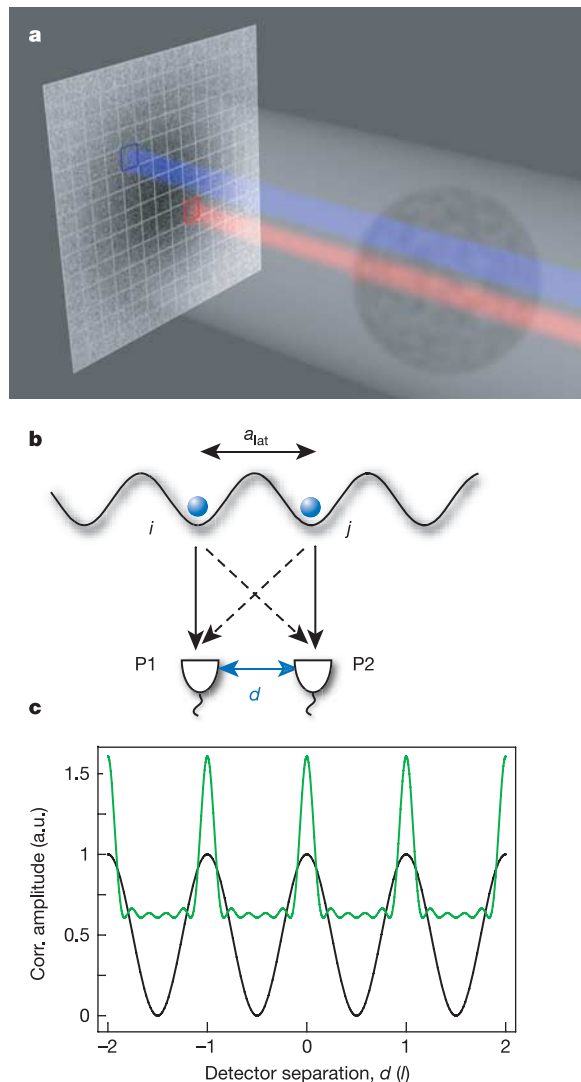
Institut für Physik, Johannes Gutenberg-Universität, Staudingerweg 7, D-55099 Mainz, Germany

In a pioneering experiment<sup>1</sup>, Hanbury Brown and Twiss (HBT) demonstrated that noise correlations could be used to probe the properties of a (bosonic) particle source through quantum statistics; the effect relies on quantum interference between possible detection paths for two indistinguishable particles. HBT correlations—together with their fermionic counterparts<sup>2–4</sup>—find numerous applications, ranging from quantum optics<sup>5</sup> to nuclear and elementary particle physics<sup>6</sup>. Spatial HBT interferometry has been suggested<sup>7</sup> as a means to probe hidden order in strongly correlated phases of ultracold atoms. Here we report such a measurement on the Mott insulator<sup>8–10</sup> phase of a rubidium Bose gas as it is released from an optical lattice trap. We show that strong periodic quantum correlations exist between density fluctuations in the expanding atom cloud. These spatial correlations reflect the underlying ordering in the lattice, and find a natural interpretation in terms of a multiple-wave HBT interference effect. The method should provide a useful tool for identifying complex quantum phases of ultracold bosonic and fermionic atoms<sup>11–15</sup>.

Although quantum noise correlation analysis is now a basic tool in various areas of physics, applications to the field of cold atoms have been scarce. Most of these concentrate on photon correlation techniques from quantum optics<sup>5,16</sup>. It was not until 1996 that bunching of cold (but non-degenerate) bosonic atom clouds could be directly measured<sup>17</sup>, followed by the observation of reduced inelastic losses due to a modification of local few-body correlations by quantum degeneracy<sup>18–20</sup>.

In our experiment, we directly measure the spatial correlation function of the density fluctuations in a freely expanding atomic

cloud<sup>7,21–23</sup>. We create an ultracold Bose gas in an optical lattice with several hundred thousand occupied lattice sites, and record the density distribution after sudden switch-off of the trapping potential and a fixed period of free expansion (the ‘time of flight’). Resonant absorption of a probe laser<sup>24</sup> yields the two-dimensional column density of the cloud, that is, the density profile integrated along the probe line of sight, as illustrated in Fig. 1a. It should be noted that the density after the time of flight reflects the in-trap momentum distribution rather than the in-trap density distribution. We performed the experiment with a Bose gas initially in the Mott insulator regime<sup>8–10</sup>, where repulsive interactions pin the atomic density to exactly an integer number of atoms per lattice site, typically between one and three. In this Mott insulator phase, the average density distribution after expansion is simply given by the



**Figure 1** Illustration of the atom detection scheme and the origin of quantum correlations. **a**, The cloud of atoms is imaged to a detector plane and sampled by the pixels of a CCD camera. Two pixels P1 and P2 are highlighted, each of which registers the atoms in a column along its line of sight. Depending on their spatial separation  $d$ , their signals show correlated quantum fluctuations, as illustrated in **b**. **b**, When two atoms initially trapped at lattice sites  $i$  and  $j$  (separated by the lattice spacing  $a_{\text{lat}}$ ) are released and detected independently at P1 and P2, the two indistinguishable quantum mechanical paths, illustrated as solid and dashed lines, interfere constructively for bosons (or destructively for fermions). **c**, The resulting joint detection probability (correlation amplitude) of simultaneously finding an atom at each detector is modulated sinusoidally as a function of  $d$  (black curve). The multiple wave generalization to a regular array of six sources with the same spacing is shown in green. a.u., arbitrary units.

incoherent sum of all single particle wavefunctions released from each lattice site—a featureless gaussian. However, a typical single shot absorption image as shown in Fig. 2a and b exhibits large fluctuations around this average. It is the purpose of this Letter to demonstrate that these fluctuations are related to intrinsic quantum noise and that their HBT-type correlations contain information on the spatial order in the lattice that is absent from the average density.

To analyse the fluctuations, we introduce the spatially averaged, normalized density–density correlation function:

$$C(\mathbf{d}) = \frac{\int \langle n(\mathbf{x} + \mathbf{d}/2) \cdot n(\mathbf{x} - \mathbf{d}/2) \rangle d^2\mathbf{x}}{\int \langle n(\mathbf{x} + \mathbf{d}/2) \rangle \langle n(\mathbf{x} - \mathbf{d}/2) \rangle d^2\mathbf{x}} \quad (1)$$

which denotes the conditional probability of finding two particles at two positions separated by a vector  $\mathbf{d}$ , averaged over all such positions. In equation (1),  $n(\mathbf{x})$  is the column density obtained from a single absorption image and the brackets  $\langle \rangle$  denote averaging over an ensemble of independently acquired images. Uncorrelated particles correspond to  $C(\mathbf{d}) = 1$ , whereas  $C(0) > 1$  indicates a tendency of particles to bunch, typical for bosons. In Fig. 2c and d, an experimentally obtained correlation function is shown. In striking contrast to the atomic density distribution of Fig. 2a, sharp peaks emerge. They appear on positions corresponding to the reciprocal lattice vectors of the original periodic trapping potential.

We found the correlation patterns to be robust in the Mott insulating regime, and observed them over a broad range of lattice depths. For a planar lattice of several thousand one-dimensional decoupled Bose gases with random phases<sup>10,25</sup>, we observed similar density correlations. The latter case is related to a recent experiment<sup>26</sup>, where single shot interference patterns were observed from 30 independent Bose–Einstein condensates with random phases. Both these cases can be described through a classical field model, whereas the case of a Mott insulator presented here requires a full quantum treatment and detection of the atom number distribution at the atomic shot noise level.

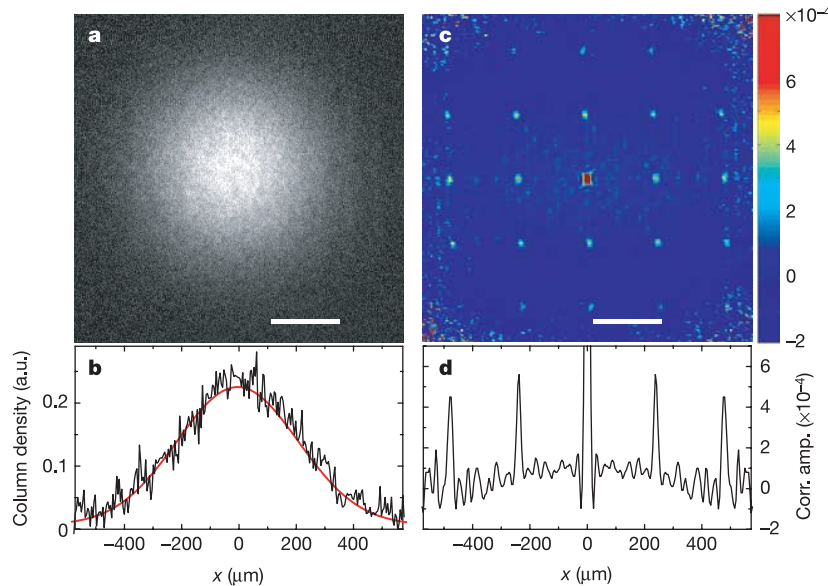
In order to explain the origin of the correlations in the density fluctuations and their regularity, let us first consider the simple

model illustrated in Fig. 1b. Two bosonic atoms in a periodic potential are initially localized at two lattice sites  $i$  and  $j$  separated by  $n_{ij}$  lattice spacings. When these particles are released from the trapping potential, one can show<sup>7,21</sup> that the joint detection probability by two detectors separated by a distance  $d$  is sinusoidally varying as a function of  $d$  (Fig. 1c). The spatial wavevector  $2\pi n_{ij}/l$  of this modulation is determined by the separation of the sources and the characteristic length:

$$l = \frac{h}{ma_{\text{lat}}} t \quad (2)$$

where  $t$  denotes the time of flight,  $h$  is Planck’s constant,  $m$  is the atomic mass, and  $a_{\text{lat}}$  is the lattice spacing. For a large number of perfectly distributed but independent sources corresponding to the individual sites of the Mott insulator, the joint detection amplitudes for all possible pairs in the source have to be added. The wave-numbers associated with every pair formed from such a distribution are then integer multiples of  $2\pi/l$ , so that away from  $d = 0$  the amplitudes add up constructively wherever the detector separation is a multiple of  $l$ . A lattice pattern of sharp peaks reproducing the reciprocal lattice therefore emerges as the number of particles increases, each peak width being roughly determined by  $l/N_s$ , with  $N_s$  being the number of occupied sites in the lattice in one dimension. Owing to the different quantum statistics of bosons and fermions the fundamental sine components have opposite signs, and therefore positive peaks emerge for bosons and negative peaks for fermions.

In the experiments, each detector in the above model is represented by a pixel of the CCD (charge-coupled device) camera, which detects the absorption image of the atom cloud. The array of pixels in this camera thereby samples the column density of the atomic cloud, according to  $n = N_{\text{bin}}/A_{\text{px}}$ . Here  $N_{\text{bin}}$  is the number of atoms detected within a column defined by the area  $A_{\text{px}}$  of each pixel and the direction of propagation of the probe light. In addition, each bin is smoothed by the point spread function of our imaging system, which we approximate by a gaussian. In our case, its root mean square (r.m.s.) radius  $\sigma \approx 5.6 \mu\text{m}$  is larger than the pixel size



**Figure 2** Single shot absorption image including quantum fluctuations and the associated spatial correlation function. **a**, Two-dimensional column density distribution of a Mott insulating atomic cloud containing  $6 \times 10^5$  atoms, released from a three-dimensional optical lattice potential with a lattice depth of  $50E_r$ . The white bars indicate the reciprocal lattice scale  $l$  defined in equation (2). **b**, Horizontal section (black line) through the centre of the image in **a**, and gaussian fit (red line) to the average over 43 independent images,

each one similar to **a**. **c**, Spatial noise correlation function obtained by analysing the same set of images, which shows a regular pattern revealing the lattice order of the particles in the trap. **d**, Horizontal profile through the centre of the pattern, containing the peaks separated by integer multiples of  $l$ . The width of the individual peaks is determined by the optical resolution of our imaging system.

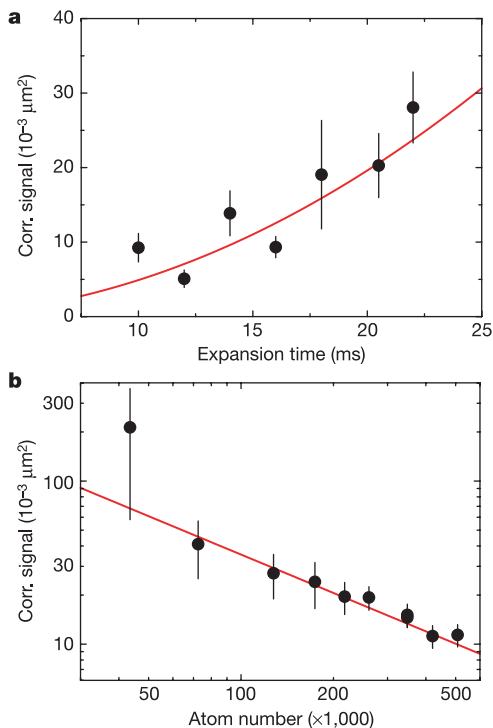
(4.4  $\mu\text{m}$ ) and determines an effective width for each bin.

The correlation for such a coarse-grained density can still be calculated from equation (1) by taking the integration over the probe line of sight into account and convoluting the resulting correlation signal with the resolution function described above (see also Methods section). The outcome of this calculation can be qualitatively understood from statistical considerations. One expects the amount of relative number fluctuations in a single detection bin to scale as  $1/\sqrt{N_{\text{bin}}}$  for  $N_{\text{bin}} \gg 1$ , implying a  $1/N_{\text{bin}}$  scaling for the correlation amplitude increase above the uncorrelated level  $C(\mathbf{d}) = 1$  (see equation (1)). However, in our case this signal is distributed over  $N_{\text{peaks}} \approx 4\pi(w/l)^2$  correlation peaks within the expanding density envelope of width  $w$ . To obtain the magnitude of the signal, we calculate the average number of atoms per bin as  $N_{\text{bin}} = N_{\text{atoms}}(\sigma/w)^2$ , with  $N_{\text{atoms}}$  representing the total atom number. This yields a correlation amplitude per peak:

$$C(\mathbf{d}_{\text{peak}}) - 1 \approx \frac{1}{N_{\text{peaks}}} \frac{1}{N_{\text{bin}}} \approx \frac{1}{4\pi N_{\text{atoms}}} \left(\frac{l}{\sigma}\right)^2 \quad (3)$$

The estimate in equation (3) agrees with a rigorous calculation assuming a homogeneous Mott insulator with unity filling (see Methods). For typical parameters of the experiments ( $5 \times 10^5$  atoms,  $l/\sigma = 40$ ), it yields a correlation amplitude of  $\sim 3 \times 10^{-4}$ , in agreement with our observations.

To confirm our analysis, we plot in Fig. 3 the experimental correlation signal from the Mott insulator versus expansion time and atom number. This signal is defined by the volume under the lateral peaks, that is, the product of the peak height times its area as determined by a gaussian fit. The resulting 'correlation signal' does not require a precise determination of the resolution, and is rather insensitive to defocusing or calibration errors. According to equation (3), it depends quadratically on the time of flight and inversely on atom number for homogeneous filling. However, for a



**Figure 3** Correlation signal versus expansion time and atom number. **a**, **b**, Average correlation signal as a function of the time of ballistic expansion (**a**) and the number of atoms  $N$  loaded in the lattice (**b**). The solid lines denote the result of a simultaneous fit to both data sets to determine the amplitude of the signal and the power law of the decay in **b** (error bars denote root-mean-square deviations).

Mott insulator in a harmonic trap, a shell structure develops for increasing filling<sup>8,27</sup>. A model of this atom number distribution (see Methods) predicts a reduction of the  $(N_{\text{atoms}})^{-1}$  scaling behaviour for large atom numbers to  $(N_{\text{atoms}})^{-0.64}$ . Using a combined fit to both data sets in Fig. 3a and b, the measured exponent of atom number scaling is  $0.78 \pm 0.15$ , close to the expected value. However, the amplitude of the signal is 40% lower than what would be expected from our simple theoretical model.

For a Bose–Einstein condensate in an optical lattice, a flat spatial correlation function is expected<sup>28</sup>. Obtaining the correlation function in this regime, however, turned out not to be experimentally possible. In the actual experiment small fluctuations of the superfluid interference pattern between the individual images exist, owing to technical reasons such as shot-to-shot atom number variations, or excitations of the condensate by external perturbations. Such fluctuations are not cancelled by the normalization in equation (1), and the associated correlations turn out to be stronger than the quantum noise correlations observed in the Mott insulating case. We attribute the much more robust quantum noise correlation signal of a Mott insulator to the gapped excitation spectrum (with an energy gap  $\sim 3$  kHz in our case), which protects this state from external perturbations that would otherwise degrade the correlations. We also investigated the case of a thermal cloud significantly above condensation temperature but with no observable population in the excited Bloch bands, for which we did not find a correlation signal. A possible explanation for this could be the decrease in signal (compared to the Mott insulating state) due to an increased spatial size of the system in combination with an increased noise background due to density and temperature fluctuations. We have considered the possibility that the correlation signal we observed in the Mott insulating case could be produced by shot-to-shot fluctuations of a residual fraction of atoms with long range coherence. In order to rule out this effect, we have checked that the regions that would contain the peaks of the diffraction pattern can be excluded from the analysis without significantly affecting the noise correlation pattern.

In conclusion, we have demonstrated that spatial quantum noise correlations in expanding atom clouds can be used to reveal the ordering of indistinguishable particles in optical lattices. They enable the direct and easy detection of many of the more complex and intriguing quantum phases that have been predicted for ultracold bosonic and fermionic atoms—for example, antiferromagnets or spin-waves in two-component spin mixtures loaded into the optical lattice<sup>12,13</sup>. Antiferromagnetic ordering or charge density waves in Fermi gases or Bose–Fermi mixtures<sup>14,15</sup>, for example, would yield additional correlation peaks at momenta given by half the reciprocal lattice vector<sup>7</sup>.

We note that after submission of this manuscript, we received a preprint<sup>29</sup> reporting the use of noise correlations in expanded atom clouds for identifying the fragments produced by ultracold molecule dissociation. □

## Methods

### Experimental sequence

Atomic Mott insulators are prepared by loading a Bose–Einstein condensate of up to  $6 \times 10^5$  atoms of  $^{87}\text{Rb}$  into an optical lattice potential. For this, three optical standing waves of wavelength  $\lambda = 850$  nm are superimposed at the position of the Bose–Einstein condensate formed in a magnetic trap. This yields a lattice of simple cubic geometry with a lattice constant of  $a_{\text{lat}} = \lambda/2 = 425$  nm. After a slow ramp-up of the lattice in 160 ms, the atoms are strongly confined at the lattice sites (potential depth  $\sim 50E_{\text{rec}}$ , with  $E_{\text{rec}} = h^2/2m\lambda^2$ ) until they are released to free ballistic expansion by switching off all potentials.

Following a time of flight period, the two-dimensional density profile of the cloud is obtained by illuminating it with a resonant laser pulse and projecting the profile of the resulting beam onto a CCD camera. A second image is taken without the atoms in the beam, and the two resulting images are divided to determine the optical density distribution of the cloud. The number of atoms in a column corresponding to a region of the imaging plane can then be deduced from the integrated optical density in that region<sup>24</sup>.



## Analysis of images

In addition to the finite resolution, the camera system adds artefacts to the images owing to optical interference effects of the coherent illumination light and electronic crosstalk during readout of the CCD chip. In our case, the former results in a pattern of vertical stripes and the latter mainly creates a periodic noise with a wavelength of two pixels. As the phase and amplitude of both periodic distortions are not constant, they can not be cancelled by the normalization procedures and appear as periodic fluctuation in the noise correlation plot. Images with high amplitude of such fluctuations (visible outside the atom cloud) are removed from further analysis. The electronic noise is addressed after the determination of the correlation function, by convolving it with a horizontal three-pixel-wide gaussian mask for smoothing.

The correlation function as defined in equation (1) is obtained from a set of images as follows: from each image the autocorrelation function (ACF) is calculated by Fourier-transforming it, taking the absolute square to obtain the power spectral density and Fourier-transforming it back. Averaging the ACF of all images yields the numerator of equation (1), whereas the denominator is obtained by calculating the ACF of the average of all images.

## Theoretical model

The origin of the correlation peaks can be understood as follows. Calculating the ACF determines the expectation value of the operator  $\langle \hat{n}(\mathbf{x}_1, t) \hat{n}(\mathbf{x}_2, t) \rangle = \langle \hat{a}^\dagger(\mathbf{x}_1, t) \hat{a}(\mathbf{x}_1, t) \hat{a}^\dagger(\mathbf{x}_2, t) \hat{a}(\mathbf{x}_2, t) \rangle$  at time  $t$ , with  $\mathbf{x}_1 = \mathbf{x} - \frac{1}{2}\mathbf{d}$ ,  $\mathbf{x}_2 = \mathbf{x} + \frac{1}{2}\mathbf{d}$ . The operators  $\hat{a}(\mathbf{x}, t)$  at position  $\mathbf{x}$  and time  $t$  after release relate to the on-site operators  $\hat{a}(\mathbf{r}_j)$  for the lattice sites  $j$  at positions  $\mathbf{r}_j$  as

$$\hat{a}(\mathbf{x}, t) = \sum_j w(\mathbf{x} - \mathbf{r}_j, t) e^{i(m/\hbar t)(\mathbf{x} - \mathbf{r}_j)^2} \hat{a}(\mathbf{r}_j)$$

where  $w$  is the expanding wavefunction originally localized to the Wannier function at the site. For the product of Fock states representing the Mott insulator with site occupation  $n_i$  at site  $i$ , one finds

$$\langle \hat{a}^\dagger(\mathbf{r}_k) \hat{a}^\dagger(\mathbf{r}_m) \hat{a}(\mathbf{r}_l) \hat{a}(\mathbf{r}_n) \rangle = n_k n_m \delta_{kl} \delta_{mn} + n_k n_m \delta_{kn} \delta_{lm} \quad (4)$$

where the delta-term introduced through the normal ordering of the operators has been omitted. In the correlation function  $C$ , the first term in equation (4) will create a constant offset of 1 for large atom number  $N$ , whereas the second term introduces a spatial dependence in the correlations, leading to:

$$C_{3D}(\mathbf{d}) = C(\mathbf{x}_1 - \mathbf{x}_2) = 1 + \frac{1}{N^2} \sum_{k,l} e^{i(m/\hbar t)(\mathbf{x}_1 - \mathbf{x}_2) \cdot (\mathbf{r}_k - \mathbf{r}_l)} n_k n_l \quad (5)$$

Throughout the discussion, constant offsets of order  $1/N$  are neglected compared to 1. For a regular one-dimensional lattice with unity filling and spacing  $a_{\text{lat}}$ , the sum can then be simplified to  $1 + \{[\sin^2(\pi Nd/l)]/[N^2 \sin^2(\pi d/l)]\}$ , with  $d = x_2 - x_1$  and  $l = \hbar t/(ma_{\text{lat}})$ , analogous to the optical interference created by a regular grating. In the limit of large  $N$ , this term corresponds to a series of peaks of height 1 and width  $l/N$  and converges to:

$$1 + \frac{1}{N} \sum_{j=-\infty}^{\infty} \delta(d/l - j)$$

For a regular three-dimensional system the structure term converges to:

$$C_{3D}(\mathbf{d}) = 1 + \frac{1}{N} \sum_j \delta((\mathbf{d} - \mathbf{p}_j) \cdot \frac{\mathbf{t}}{m})/l$$

where  $\mathbf{p}_j$  are the reciprocal three-dimensional lattice momenta. Because the imaging system registers only column densities and has a finite resolution, the operators  $\hat{n}(\mathbf{x}_{1,2})$  both have to be convolved with the inverse point spread function (approximated as a gaussian of r.m.s. width  $\sigma$ ) and integrated along the imaging axis before being evaluated. For unity filling this yields a smoothed two-dimensional correlation function:

$$C(\mathbf{d}) = 1 + \frac{1}{4\pi N} \left(\frac{l}{\sigma}\right)^2 \sum_j e^{-[|\mathbf{d} - \mathbf{p}_j|/m]^2/4\sigma^2}$$

The heights of the peaks at the reciprocal lattice momenta therefore scale as  $N^{-1}l^2$  for this simple homogeneous case. As indicated in the text, the  $N^{-1}$  scaling is modified to  $N^{-0.64}$  for our harmonically trapped system by the appearance of Mott domains with filling factor larger than one for higher atom numbers. The prediction for the exponent has been obtained by numerically evaluating the sum in equation (5) using a model distribution of atoms in the lattice sites confined by a global parabolic potential. This distribution is predicted assuming the system can be described in the strongly interacting limit<sup>30</sup> with a local density approximation.

Received 27 January; accepted 28 February 2005; doi:10.1038/nature03500.

- Hanbury Brown, R. & Twiss, R. Q. Correlation between photons in two coherent beams of light. *Nature* **177**, 27–29 (1956).
- Henny, M. *et al.* The fermionic Hanbury Brown and Twiss experiment. *Science* **284**, 296–298 (1999).
- Oliver, W. D., Kim, J., Liu, R. C. & Yamamoto, Y. Hanbury Brown and Twiss-type experiment with electrons. *Science* **284**, 299–301 (1999).
- Kiesel, H., Renz, A. & Hasselbach, F. Observation of Hanbury Brown-Twiss anticorrelations for free electrons. *Nature* **418**, 392–394 (2002).
- Bachor, H. A. & Ralph, T. C. A *Guide to Experiments in Quantum Optics* (Wiley-VCH, Weinheim, 2004).
- Baym, G. The physics of Hanbury Brown-Twiss intensity interferometry: From stars to nuclear collisions. *Act. Phys. Pol. B* **29**, 1839–1884 (1998).
- Altman, E., Demler, E. & Lukin, M. D. Probing many-body states of ultracold atoms via noise correlations. *Phys. Rev. A* **70**, 013603 (2004).
- Jaksch, D., Bruder, C., Cirac, J. I., Gardiner, C. W. & Zoller, P. Cold bosonic atoms in optical lattices. *Phys. Rev. Lett.* **81**, 3108–3111 (1998).

- Greiner, M., Mandel, O., Esslinger, T., Hänsch, T. W. & Bloch, I. Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms. *Nature* **415**, 39–44 (2002).
- Stöferle, T., Moritz, H., Schori, C., Köhl, M. & Esslinger, T. Transition from a strongly interacting 1D superfluid to a Mott insulator. *Phys. Rev. Lett.* **92**, 130403 (2004).
- Hofstetter, W., Cirac, J. I., Zoller, P., Demler, E. & Lukin, M. D. High temperature superfluidity of fermionic atoms in optical lattices. *Phys. Rev. Lett.* **89**, 220407 (2002).
- Kuklov, A. & Svistunov, B. Counterflow superfluidity of two-species ultracold atoms in a commensurate optical lattice. *Phys. Rev. Lett.* **90**, 100401 (2003).
- Duan, L.-M., Demler, E. & Lukin, M. Controlling spin exchange interactions of ultracold atoms in an optical lattice. *Phys. Rev. Lett.* **91**, 090402 (2003).
- Lewenstein, M., Santos, L., Baranov, M. A. & Fehrmann, H. Atomic Bose-Fermi mixtures in an optical lattice. *Phys. Rev. Lett.* **92**, 050401 (2004).
- Roth, R. & Burnett, K. Quantum phases of atomic boson-fermion mixtures in optical lattices. *Phys. Rev. A* **69**, 021601(R) (2004).
- Jurczak, C. *et al.* Atomic transport in an optical lattice: An investigation through polarization-selective intensity correlations. *Phys. Rev. Lett.* **77**, 1727–1730 (1996).
- Yasuda, M. & Shimizu, F. Observation of two-atom correlation of an ultracold neon atomic beam. *Phys. Rev. Lett.* **77**, 3090–3093 (1996).
- Kagan, Y., Svistunov, B. V. & Shlyapnikov, G. V. Effect of Bose condensation on inelastic processes in gases. *Sov. Phys. JETP Lett.* **42**, 209–212 (1985).
- Burt, E. A. *et al.* Coherence, correlations, and collisions: What one learns about Bose-Einstein condensates from their decay. *Phys. Rev. Lett.* **79**, 337–340 (1997).
- Laburthe-Tolra, B. *et al.* Observation of reduced three-body recombination in a correlated 1D degenerate Bose gas. *Phys. Rev. Lett.* **92**, 190401 (2004).
- Grondalski, J., Alsing, P. M. & Deutsch, I. H. Spatial correlation diagnostics for atoms in optical lattices. *Opt. Exp.* **5**, 249–261 (1999).
- Kolovsky, A. R. Interference of cold atoms released from an optical lattice. *Europhys. Lett.* **68**, 330–336 (2004).
- Bach, R. & Rzaewski, K. Correlation functions of cold bosons in an optical lattice. *Phys. Rev. A* **70**, 063622 (2004).
- Ketterle, W., Durfee, D. S. & Stamper-Kurn, D. M. in *Proc. Int. School of Physics “Enrico Fermi”* (eds Inguscio, M., Stringari, S. & Wieman, C. E.) 67–176 (IOS Press, Amsterdam, 1999).
- Greiner, M., Bloch, I., Mandel, O., Hänsch, T. W. & Esslinger, T. Exploring phase coherence in a 2D lattice of Bose-Einstein condensates. *Phys. Rev. Lett.* **87**, 160405 (2001).
- Hadzibabic, Z., Stock, S., Battelier, B., Bretin, V. & Dalibard, J. Interference of an array of independent Bose-Einstein condensates. *Phys. Rev. Lett.* **93**, 180403 (2004).
- Batrouni, G. G. *et al.* Mott domains of bosons confined on optical lattices. *Phys. Rev. Lett.* **89**, 117203 (2003).
- Naraszewski, M. & Glauber, R. Spatial coherence and density correlations of trapped Bose gases. *Phys. Rev. A* **59**, 4595–4607 (1999).
- Greiner, M., Regal, C. A., Stewart, J. T. & Jin, D. S. Probing pair-correlated fermionic atoms through correlations in atom shot noise. Preprint at (<http://arxiv.org/cond-mat/0502411>) (2005).
- Sheshadri, K., Krishnamurthy, H. R., Pandit, R. & Ramakrishnan, T. V. Superfluid and insulating phases in an interacting-boson model: mean-field theory and the RPA. *Europhys. Lett.* **22**, 257–263 (1993).

**Acknowledgements** We acknowledge discussions with E. Altman and M. Greiner, as well as financial support by the DFG, AFOSR and the EU under a Marie-Curie Fellowship (E.G.) and a Marie-Curie Excellence grant.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to I.B. (bloch@uni-mainz.de).

## Orbital Kondo effect in carbon nanotubes

Pablo Jarillo-Herrero, Jing Kong\*, Herre S.J. van der Zant, Cees Dekker, Leo P. Kouwenhoven & Silvano De Franceschi\*

Kavli Institute of Nanoscience, Delft University of Technology, PO Box 5046, 2600 GA Delft, The Netherlands

\* Present addresses: Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA (J.K.); Laboratorio Nazionale TASC-INFN, I-34012 Trieste, Italy (S.D.F.)

**Progress in the fabrication of nanometre-scale electronic devices is opening new opportunities to uncover deeper aspects of the Kondo effect<sup>1</sup>—a characteristic phenomenon in the physics of strongly correlated electrons. Artificial single-impurity Kondo systems have been realized in various nanostructures, including semiconductor quantum dots<sup>2–4</sup>, carbon nanotubes<sup>5,6</sup> and individual molecules<sup>7,8</sup>. The Kondo effect is usually regarded as a**

spin-related phenomenon, namely the coherent exchange of the spin between a localized state and a Fermi sea of delocalized electrons. In principle, however, the role of the spin could be replaced by other degrees of freedom, such as an orbital quantum number<sup>9,10</sup>. Here we show that the unique electronic structure of carbon nanotubes enables the observation of a purely orbital Kondo effect. We use a magnetic field to tune spin-polarized states into orbital degeneracy and conclude that the orbital quantum number is conserved during tunnelling. When orbital and spin degeneracies are present simultaneously, we observe a strongly enhanced Kondo effect, with a multiple splitting of the Kondo resonance at finite field and predicted to obey a so-called SU(4) symmetry.

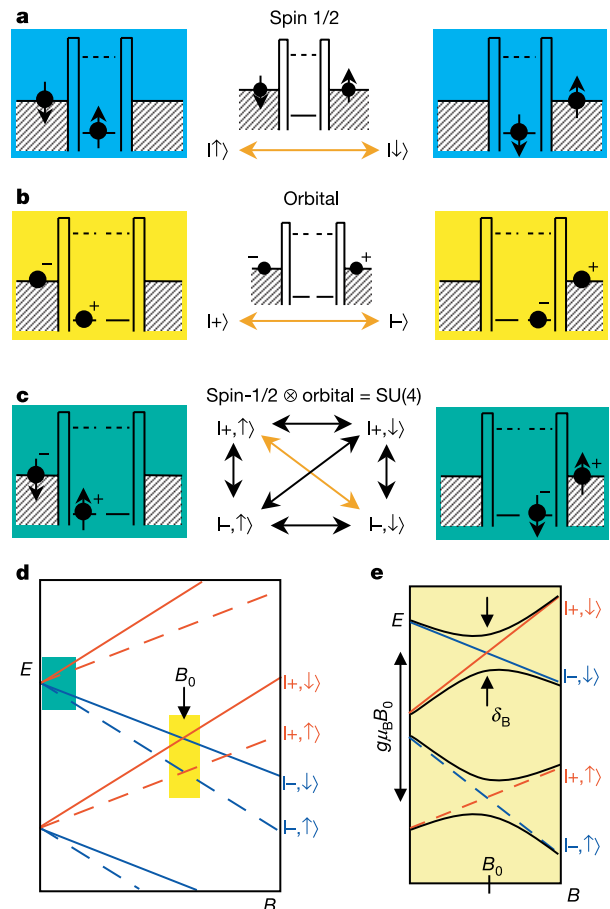
The simplest Kondo system consists of a localized, spin- $\frac{1}{2}$  electron coupled to a Fermi sea by means of a Heisenberg-like exchange interaction<sup>1</sup>. This simple system can be realized with a quantum dot (QD) device<sup>2–4</sup>, which is a small electronic island connected to metallic leads via two tunnel barriers (see Fig. 1a). Below a characteristic temperature  $T_K$ , the so-called Kondo temperature, a many-body singlet state is formed between the QD spin and the surrounding conduction electrons (Fig. 1a). This state adds a resonant level at the Fermi energy of the electrodes, enabling the tunnelling of electrons across the QD. Such a Kondo resonance can lead to a strong enhancement of the conductance, overcoming the Coulomb blockade effect<sup>2–4</sup>. In principle, a Kondo effect might also occur in the absence of spin if another quantum number, for example an orbital degree of freedom, gives rise to a degeneracy (Fig. 1b). In this case, Kondo correlations lead to the screening of the local orbital ‘polarization’, and an orbital singlet is formed through a combination of orbital states. In the presence of both spin and orbital degeneracy, quantum fluctuations lead to a strong mixing of these two degrees of freedom (Fig. 1c). This increased degeneracy yields an enhancement of  $T_K$  (ref. 11). In the low-temperature limit, this system is described by a Hamiltonian obeying SU(4) symmetry, which implies that the spin and charge degrees of freedom are fully entangled and the state of the electron is represented by a four-component ‘hyperspin’<sup>12–15</sup>.

An orbital degeneracy is naturally expected in the electronic structure of carbon nanotubes<sup>16</sup> (CNTs). This degeneracy can intuitively be viewed to originate from the two equivalent ways in which electrons can circle around the graphene cylinder, namely clockwise and anticlockwise<sup>17</sup>. The rotational motion confers an orbital magnetic moment on the electrons. Consequently, the orbital degeneracy can be split by a magnetic field,  $B$ , parallel to the nanotube axis. (Experimental evidence for this effect, originally predicted by Ajiki and Ando<sup>18</sup>, has recently been reported<sup>17,19–21</sup>.) We label the orbital states of a CNT QD as  $|+\rangle$  and  $|-\rangle$  according to the sign of the energy shift that they experience under an applied  $B$ . Size quantization due to the finite CNT length results in two sets of orbital levels,  $E_+^{(n)}$  and  $E_-^{(n)}$ , where  $n = 1, 2, 3, \dots$  is the quantization number.  $E_+^{(n)} = E_-^{(n)}$  at  $B = 0$  (assuming no orbital mixing), resulting in a fourfold degeneracy when including spin. The orbital and spin degeneracies are simultaneously lifted by a parallel  $B$  (Fig. 1d). The use of  $B$  permits the tuning of new degeneracies in connection with the crossing between levels from different shells. Here we are particularly interested in the crossing between states with the same spin polarization, of the type indicated by the yellow rectangle in Fig. 1d. We show below that the twofold orbital degeneracy originating from such a crossing gives rise to a purely orbital Kondo effect. We then consider the case of concomitant spin and orbital degeneracy at  $B = 0$  (green rectangle in Fig. 1d) and present evidence for an SU(4) Kondo effect.

In a measurement of the linear conductance,  $G$ , as a function of gate voltage,  $V_G$ , the fourfold shell structure leads to consecutive groups of four closely spaced Coulomb blockade oscillations<sup>6,22</sup>. The  $B$ -evolution of such oscillations is shown in Fig. 2a for a CNT QD device (described in the inset to Fig. 2a and in the figure legend) in a

$V_G$  region encompassing two adjacent shells. Coulomb peaks (highlighted by green lines) appear as lines running from bottom to top and denote the sequential addition of electrons to the QD; the electron number increases from left to right. The observed pattern is explained in detail on the basis of the single-particle spectrum in Fig. 1d and taking into account the Coulomb interaction between electrons (see Supplementary Information).

The Coulomb peaks move to the left or right when increasing  $B$ ,



**Figure 1** Spin, orbital and SU(4) Kondo effect in a quantum dot (QD) with an odd number of electrons. The left and right panels in **a–c** represent initial and final ground states, respectively. **a**, Schematic illustration of a spin-flip co-tunnelling process connecting the two states spin up,  $|\uparrow\rangle$ , and down,  $|\downarrow\rangle$ , from a single orbital state. The intermediate virtual state is shown in the central diagram. This co-tunnelling event is one of many higher-order processes that add up coherently, resulting in the screening of the local spin. **b**, Co-tunnelling process for spinless electrons for two degenerate orbital states, labelled  $|+\rangle$  and  $|-\rangle$ . The depicted process flips the orbital quantum number from  $|+\rangle$  to  $|-\rangle$  and vice versa. The coherent superposition of orbital-flip processes leads to the screening of the local orbital quantum number. **c**, QD with two spin-degenerate orbitals leading to an overall fourfold degeneracy. Spin and/or orbital states can flip by one-step co-tunnelling processes, indicated by black arrows in the central diagram; the orange arrow refers to the co-tunnelling event connecting the two states depicted in the green diagrams. These processes lead to the entanglement of spin and orbital states, resulting in an enhanced SU(4) Kondo effect. **d**, Qualitative single-particle energy spectrum of a CNT QD in a magnetic field. Red and blue lines represent, respectively, orbital states shifting up,  $|+\rangle$ , and down,  $|-\rangle$ , in energy. Dashed and solid lines represent spin-up and spin-down states, respectively. The yellow rectangle highlights the region where a purely orbital Kondo effect can occur because of a level-crossing (at  $B = B_0$ ) between spin-polarized states. The green rectangle highlights the SU(4) Kondo region. **e**, Close-up of the yellow rectangle in **d**. A finite coupling,  $\delta_B$ , between  $|+\rangle$  and  $|-\rangle$  states causes an anticrossing (black lines). At high  $B$ ,  $\delta_B$  is smaller than the Zeeman splitting,  $g\mu_B B$ .

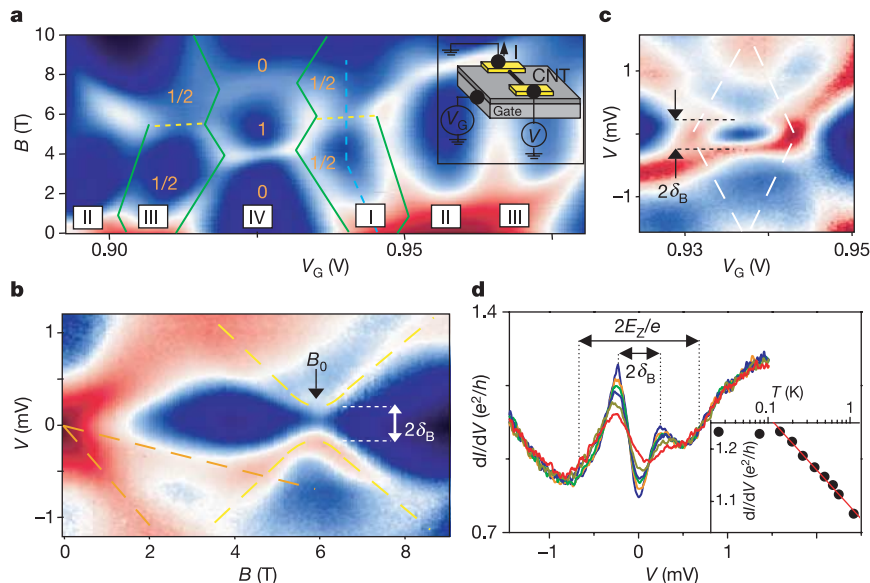
which corresponds to adding the last electron to a  $|-\rangle$  or  $|+\rangle$  orbital, respectively. When the ground-state configuration of the QD changes, kinks appear in the  $B$ -evolution of the Coulomb peaks. The two enhanced-conductance ridges at  $B = B_0 \sim 6$  T, bounded by two such kinks and highlighted by dotted yellow lines, are due to the crossing between  $|-\rangle$  and  $|+\rangle$  states as described in Fig. 1d. A detailed analysis (see Supplementary Information) indicates that along these ridges the QD ground state is doubly degenerate, with the last added electron occupying the level crossing between  $|+, \uparrow\rangle$  and  $|-, \uparrow\rangle$  (left ridge) or between  $|+, \downarrow\rangle$  and  $|-, \downarrow\rangle$  (right ridge).

In the region near the degeneracy point, we are able to measure a small coupling between orbital states<sup>21</sup>, resulting in level repulsion at  $B = B_0$ . The energy splitting is directly observed in the spectroscopy data of Fig. 2b, where the differential conductance,  $dI/dV$ , is shown as a function of  $B$  and bias voltage,  $V$ . In this measurement  $V_G$  and  $B$  are simultaneously varied to follow the middle of the Coulomb valley (dashed blue line in Fig. 2a). Here, single-electron tunnelling is suppressed and the spectroscopy is based on higher-order co-tunnelling processes, which lead to an enhancement of  $dI/dV$  every time that  $V$  equals an internal excitation energy<sup>23</sup>. We focus on the high- $B$  region of Fig. 2b. As  $B$  is swept across  $B_0$ , the anticrossing between  $|+, \downarrow\rangle$  and  $|-, \downarrow\rangle$  (depicted in Fig. 1e) shows up in the two  $dI/dV$  ridges highlighted by dashed yellow lines. The level spacing, corresponding to half the distance between these lines, reaches a minimum value  $\delta_B = 225 \mu\text{V}$  at  $B = B_0 = 5.9$  T. In a measurement of  $dI/dV$  against  $(V, V_G)$  at 5.9 T, shown in Fig. 2c, the higher-order peaks appear as horizontal ridges inside the Coulomb diamond. Their spacing,  $2\delta_B$ , is independent of  $V_G$ , whereas their height increases towards the edges of the diamond.

An individual trace of  $dI/dV$  against  $V$  taken in the middle of the

diamond is shown in Fig. 2d, together with traces measured at higher temperature,  $T$ . The strong overshoot of the  $dI/dV$  peaks and their log- $T$  dependence (inset) indicate an important contribution from Kondo correlations. The observed behaviour is characteristic of a split Kondo resonance—that is, a Kondo resonance associated with two quasi-degenerate states—in line with recent theoretical predictions<sup>24</sup> and experiments<sup>25</sup>. It is important to note that the Zeeman spin splitting,  $E_Z = g\mu_B B_0$ , is three times  $\delta_B$ , indicating that the Kondo effect originates entirely from orbital correlations occurring at the crossing between two spin-polarized states,  $|+, \downarrow\rangle$  and  $|-, \downarrow\rangle$ . This conclusion is in agreement with the zero-field data that we show below. The large Zeeman splitting also ensures that the observed orbital Kondo resonance provides a conducting channel only for  $|\downarrow\rangle$  electrons, thereby acting as a high-transmission spin filter<sup>12–14</sup>. In contrast, the conductance enhancement that occurs for three-electron shell filling originates from  $|+, \uparrow\rangle$  and  $|-, \uparrow\rangle$  states and hence it allows only the tunnelling of  $|\uparrow\rangle$  electrons. Switching from one degeneracy to the other is controlled by simply switching the gate voltage, which then causes the CNT QD to operate as a bipolar spin filter.

We now centre our attention on the zero-field regime, in which both orbital and spin degeneracies are expected (green rectangle in Fig. 1d). The Coulomb oscillations corresponding to the filling of a single shell are shown in Fig. 3a for a different CNT QD device. The four oscillations are clearly visible at 8 K (red trace). At lower  $T$ , the conductance exhibits a pronounced enhancement in regions I and III—that is, for 1 and 3 electrons on the shell—and the corresponding Coulomb blockade valleys disappear completely at 0.3 K (black trace). This conductance enhancement is a hallmark of Kondo correlations. From the  $T$ -dependence (fully shown in the Supplementary Information) we estimate that  $T_K = 7.7$  K, an unusually high value that can be ascribed to the enhanced degeneracy<sup>11</sup>.



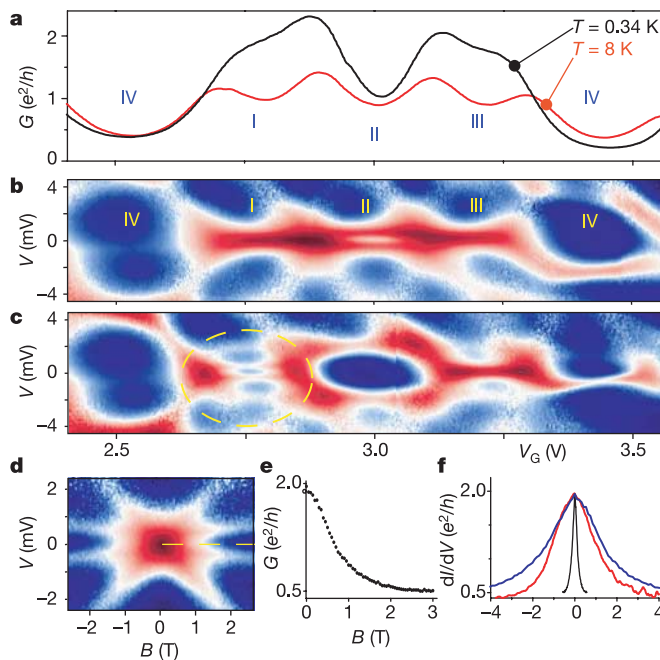
**Figure 2** Orbital Kondo effect. **a**, Colour-scale representation of the linear conductance,  $G$ , versus  $B$  and  $V_G$  at  $T \sim 30$  mK ( $G$  increases from dark blue to dark red). The green lines highlight the  $B$ -evolution of the Coulomb peaks. The dotted yellow lines highlight regions of enhanced conductance due to Kondo effect. Roman labels indicate the number of electrons on the last occupied shell near  $B = 0$ . Orange numbers indicate the spin of the ground state. Inset: device scheme. Carbon nanotubes were grown by chemical vapour deposition on p-type silicon substrates with a surface oxide 250 nm thick. Individual nanotubes were located by atomic force microscopy and contacted with Ti/Au electrodes (typical separation  $\sim 100$ – $800$  nm) defined by e-beam lithography. The highly doped silicon substrate was used as a back-gate. **b**, Colour-scale plot of the differential

conductance,  $dI/dV$ , against  $V$  and  $B$  along the dashed blue line in **a**. The field splits the Kondo resonance into multiple peaks. The two orange lines highlight the evolution of the peaks associated with the spin and orbital splitting, respectively. The spectroscopy features are more pronounced for  $V < 0$ , most probably due to asymmetric tunnel barriers<sup>30</sup>. The yellow lines highlight the orbital anticrossing at  $B = B_0 = 5.9$  T. **c**, Coulomb diamond for one electron on the last occupied shell at  $B = 5.9$  T. **d**, Plot of  $dI/dV$  against  $V$  with varying  $T$ , from 25 mK (thick blue trace) to 1.1 K (thick red trace), at the anticrossing point ( $B = 5.9$  T,  $V_G = 937$  mV). Orbital splitting,  $\delta_B$ , and Zeeman splitting,  $E_Z$ , are compared visually. The split Kondo peaks decrease with increasing  $T$ . Inset: plot of peak height against  $T$  evaluated for the left peak.



The important contribution of the orbital degree of freedom becomes apparent from the  $B$ -dependence of  $G$  (Fig. 3e, f). If this Kondo effect were determined by spin only (this could be the case if one of the orbitals was coupled weakly to the leads),  $G$  should decrease on a field scale  $B \sim k_B T_K / g \mu_B \sim 6$  T due to Zeeman splitting<sup>26</sup>. In contrast,  $G$  decays on a much smaller scale,  $B \sim k_B T_K / 2 \mu_{\text{orb}} \sim 0.5$  T, which is determined by the orbital splitting (an estimate of the orbital magnetic moment,  $\mu_{\text{orb}}$ , is given below).

In the nonlinear regime, a single zero-bias Kondo resonance appears in regions I and III (Fig. 3b). Contrary to the result in Fig. 2c, no orbital splitting is observed because of the much larger  $T_K$  ( $k_B T_K > \delta$  (refs 12, 13, 21)). In region II we observe two peaks at finite bias, reflecting the already known splitting of a singlet–triplet Kondo resonance<sup>27</sup>. To show that the Kondo resonance in regions I and III arises from simultaneous orbital and spin Kondo correlations, we investigate the effect of lifting spin and orbital degeneracies at finite  $B$ . As opposed to an ordinary spin- $\frac{1}{2}$  Kondo system (where the Kondo resonance splits in two peaks, separated by twice the Zeeman energy<sup>3–9</sup>) we find a fundamentally different splitting. At  $B = 1.5$  T (Fig. 3c), multiple split peaks appear in regions I and III as ridges of enhanced  $dI/dV$  parallel to the  $V_G$  axis. In region I,



**Figure 3** Kondo effect with combined spin and orbital degeneracies (spin $\otimes$ orbital Kondo effect). **a**, Plot of linear conductance,  $G$ , against  $V_G$  at  $T = 8$  K (red) and  $0.34$  K (black). **b**, Colour-scale plot of the differential conductance,  $dI/dV$ , against  $(V, V_G)$  at  $T = 0.34$  K and  $B = 0$  ( $dI/dV$  increases from blue to red). **c**, As in **b**, but at  $B = 1.5$  T. The circle indicates the fourfold splitting in region I. **d**, Colour-scale plot of  $dI/dV$  against  $(V, B)$  in the centre of Coulomb valley I. The Kondo peak appears as a bright spot at  $(V, B) = (0, 0)$ , and splits into four peaks at finite  $B$ , following the simultaneous splitting of the orbital and spin states. **e**,  $B$ -dependence of  $G$  taken from the zero-bias dashed yellow line in **d**.  $G$  decreases on a  $\sim 0.5$  T scale; that is,  $\sim 12$ -fold faster than expected from Zeeman splitting. **f**, Plot of  $G$  against normalized Zeeman energy,  $g\mu_B B/k_B T_K$  (black trace), and against normalized orbital splitting,  $2\mu_{\text{orb}} B/k_B T_K$  (blue trace).  $T_K = 7.7$  K as deduced from a fit of  $G(T)$ . (Note that  $G(B) = 0.5G(0)$  when  $2\mu_{\text{orb}} B/k_B T_K \sim 1$ .) To compare the suppression due to  $B$  with that due to  $V$ , we also show a measurement of  $dI/dV$  against normalized bias voltage,  $eV/k_B T_K$  (red trace). The blue and red traces fall almost on top of each other, implying that lifting the orbital degeneracy suppresses the Kondo effect. This shows that the simultaneous degeneracy of orbital and spin states forms the origin of the strongly enhanced Kondo effect at  $B = 0$ .

the large zero-bias resonance opens up in four peaks that move linearly with  $B$  and become progressively smaller (Fig. 3d). The two inner peaks are due to Zeeman splitting, that is, to higher-order co-tunnelling from  $|-, \uparrow\rangle$  to  $|-, \downarrow\rangle$  ( $|-\rangle$  is the lower-energy orbital). The two outer peaks arise from co-tunnelling from orbital  $|-\rangle$  to orbital  $|+\rangle$ . In the latter case, inter-orbital co-tunnelling processes can occur either with or without spin flip. (However, the corresponding substructure<sup>21</sup> is not resolved owing to the broadening of the outer peaks.) Similar multiple splittings of the Kondo resonance have also been observed in several other samples. According to recent calculations<sup>28</sup>, the observed multiple splitting of the Kondo resonance constitutes direct evidence of SU(4) symmetry, which implies the concomitant presence of spin as well as orbital Kondo correlations, confirming our previous finding.

The slope  $|dV/dB|$  of a conductance peak (Fig. 3d) directly yields the value of the magnetic moment associated with the splitting. We obtain a spin magnetic moment  $\mu_{\text{spin}} = \frac{1}{2}|dV/dB|_{\text{spin}} = 0.06 \text{ meV T}^{-1} \sim \mu_B$  from the inner peaks, and an orbital magnetic moment  $\mu_{\text{orb}} = \frac{1}{2}|dV/dB|_{\text{orb}}/\cos\varphi = 0.8 \text{ meV T}^{-1} \sim 13\mu_B$  from the outer peaks ( $\varphi$  is the angle between the nanotube and  $B$ )<sup>17</sup>. The same value of  $\mu_{\text{orb}}$  follows from the splitting of the Kondo resonance in region III (Fig. 3c). In this case, however, no Zeeman splitting is observed. Here, the magnetic field induces a transition from SU(4) to a spin-based SU(2) Kondo effect for which  $k_B T_K$  remains larger than the Zeeman energy, hindering the splitting of the Kondo resonance up to a few teslas. Finally, we note that both the one-electron SU(4) and the two-electron singlet–triplet Kondo effects are characterized by a fourfold degeneracy, which results in an enhanced  $T_K$  (ref. 27). Apart from this, the two phenomena are fundamentally very different. The singlet–triplet Kondo effect is a spin phenomenon in which the role of the orbital degree of freedom is simply to provide the basis for the construction of spin-singlet and triplet two-particle states (see also Supplementary Information).

Because orbital Kondo correlations can arise only if the orbital quantum number is conserved during tunnelling, our experimental finding of orbital Kondo physics in CNT QDs raises an interesting question concerning the nature of the dot–lead coupling. In our devices, the metal contacts are deposited on top of the CNT and the QD is formed in the segment between them<sup>29</sup>. It is possible that when electrons tunnel out of the QD, they enter first the nanotube section underneath the contacts, where they dwell for some time before moving into the metal. Because the orbital quantum number is probably conserved in a CNT–CNT tunnel process, this intermediate step may account for the observed orbital Kondo effect. □

Received 22 November 2004; accepted 25 January 2005; doi:10.1038/nature03422.

- Hewson, A. C. *The Kondo Problem to Heavy Fermions* (Cambridge Univ. Press, Cambridge, 1993).
- Goldhaber-Gordon, D. *et al.* Kondo effect in a single-electron transistor. *Nature* **391**, 156–159 (1998).
- Cronenwett, S. M., Oosterkamp, T. H. & Kouwenhoven, L. P. A tunable Kondo effect in quantum dots. *Science* **281**, 540–544 (1998).
- Schmid, J., Weis, J., Eberl, K. & von Klitzing, K. A quantum dot in the limit of strong coupling to reservoirs. *Physica B* **256–258**, 182–185 (1998).
- Nygård, J., Cobden, D. H. & Lindelof, P. E. Kondo physics in carbon nanotubes. *Nature* **408**, 342–346 (2000).
- Buitelaar, M. R., Bachtold, A., Nussbaumer, T., Iqbal, M. & Schönenberger, C. Multi-wall carbon nanotubes as quantum dots. *Phys. Rev. Lett.* **88**, 156801 (2002).
- Park, J. *et al.* Coulomb blockade and the Kondo effect in single-atom transistors. *Nature* **417**, 722–725 (2002).
- Liang, W., Shores, M. P., Bockrath, M., Long, J. R. & Park, H. Kondo resonance in a single-molecule transistor. *Nature* **417**, 725–729 (2002).
- Cox, D. L. & Zawadowski, A. Exotic Kondo effects in metals: magnetic ions in a crystalline electric field and tunnelling centers. *Adv. Phys.* **47**, 599–942 (1998).
- Kolesnychenko, O. Yu., de Kort, R., Katsnelson, M. I., Lichtenstein, A. I. & van Kempen, H. Real-space imaging of an orbital Kondo resonance on the Cr(001) surface. *Nature* **415**, 507–509 (2002).
- Inoshita, T., Shimizu, A., Kuramoto, Y. & Sakaki, H. Correlated electron transport through a quantum dot: the multiple-level effect. *Phys. Rev. B* **48**, 14725–14728 (1993).
- Borda, L., Zarand, G., Hofstetter, W., Halperin, B. I. & von Delft, J. SU(4) Fermi liquid state and spin filtering in a double quantum dot system. *Phys. Rev. Lett.* **90**, 026602 (2003).

13. Zaránd, G., Brataas, A. & Goldhaber-Gordon, D. Kondo effect and spin filtering in triangular artificial atoms. *Solid State Commun.* **126**, 463–466 (2003).
14. López, R. *et al.* Probing spin and orbital Kondo effects with a mesoscopic interferometer. Preprint available at (<http://xxx.lanl.gov/abs/cond-mat/0402361>).
15. Sasaki, S., Amaha, S., Asakawa, N., Eto, M. & Tarucha, S. Enhanced Kondo effect via tuned orbital degeneracy in a spin  $\frac{1}{2}$  artificial atom. *Phys. Rev. Lett.* **93**, 017205 (2004).
16. Dresselhaus, M. S., Dresselhaus, G. & Eklund, P. C. *Science of Fullerenes and Carbon Nanotubes* (Academic, San Diego, 1996).
17. Minot, E., Yaish, Y., Sazonova, V. & McEuen, P. L. Determination of electron orbital magnetic moments in carbon nanotubes. *Nature* **428**, 536–539 (2004).
18. Ajiki, H. & Ando, T. Electronic states of carbon nanotubes. *J. Phys. Soc. Jpn.* **62**, 1255–1266 (1993).
19. Zaric, S. *et al.* Optical signatures of the Aharonov–Bohm phase in single-walled carbon nanotubes. *Science* **304**, 1129–1131 (2004).
20. Coskun, U. C., Wei, T.-C., Vishveshwara, S., Goldbart, P. M. & Bezryadin, A. *h/e* magnetic flux modulation of the energy gap in nanotube quantum dots. *Science* **304**, 1132–1134 (2004).
21. Jarillo-Herrero, P. *et al.* Electronic transport spectroscopy of carbon nanotubes in a magnetic field. Preprint available at (<http://xxx.lanl.gov/cond-mat/0411440>).
22. Liang, W., Bockrath, M. & Park, H. Shell filling and exchange coupling in single-walled carbon nanotubes. *Phys. Rev. Lett.* **88**, 126801 (2002).
23. De Franceschi, S. *et al.* Electron cotunneling in a semiconductor quantum dot. *Phys. Rev. Lett.* **86**, 878–881 (2001).
24. Paaske, J., Rosch, A. & Wölfe, P. Nonequilibrium transport through a Kondo dot in a magnetic field: perturbation theory. *Phys. Rev. B* **69**, 155330 (2004).
25. Pasupathy, A. N. *et al.* The Kondo effect in the presence of ferromagnetism. *Science* **306**, 86–89 (2004).
26. Costi, T. A. Kondo effect in a magnetic field and the magnetoresistivity of Kondo alloys. *Phys. Rev. Lett.* **85**, 1504–1507 (2000).
27. Sasaki, S. *et al.* Kondo effect in an integer-spin quantum dot. *Nature* **405**, 764–767 (2000).
28. Choi, M.-S., López, R. & Aguado, R. SU(4) Kondo effect in carbon nanotubes. *Phys. Rev. Lett.* (submitted).
29. Mann, D., Javey, A., Kong, J., Wang, Q. & Dai, H. J. Ballistic transport in metallic nanotubes with reliable Pd ohmic contacts. *Nano Lett.* **3**, 1541–1544 (2003).
30. Bonet, E., Deshmukh, M. M. & Ralph, D. C. Solving rate equations for electron tunneling via discrete quantum states. *Phys. Rev. B* **65**, 045317 (2002).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank G. Zaránd, R. Aguado and J. Martinek for discussions. Financial support was obtained from the Japanese Solution Oriented Research for Science and Technology (SORST) program and the Dutch Fundamenteel Onderzoek der Materie (FOM).

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to P.J. ([pablo@qt.tn.tudelft.nl](mailto:pablo@qt.tn.tudelft.nl)).

## Compact, stable and efficient all-fibre gas cells using hollow-core photonic crystal fibres

F. Benabid, F. Couny, J. C. Knight, T. A. Birks & P. St J. Russell

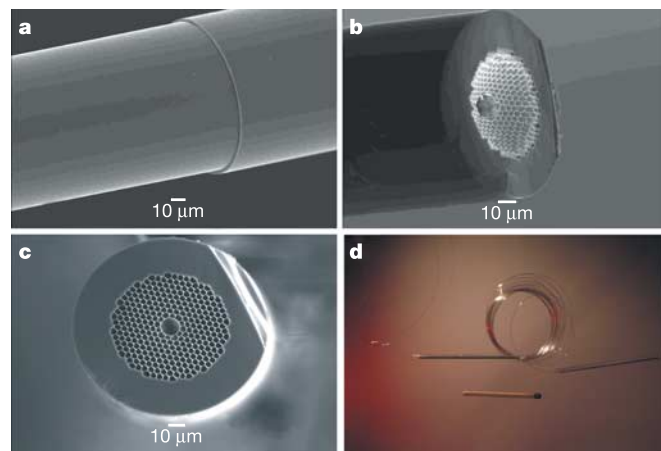
Photonics and Photonic Materials Group, Department of Physics, University of Bath, Bath BA2 7AY, UK

Gas-phase materials are used in a variety of laser-based applications—for example, in high-precision frequency measurement<sup>1,2</sup>, quantum optics and nonlinear optics<sup>3,4</sup>. Their full potential has however not been realized because of the lack of a suitable technology for creating gas cells that can guide light over long lengths in a single transverse mode while still offering a high level of integration in a practical and compact set-up or device. As a result, solid-phase materials are still often favoured, even when their performance compares unfavourably with gas-phase systems. Here we report the development of all-fibre gas cells that meet these challenges. Our structures are based on gas-filled hollow-core photonic crystal fibres, in which we have recently demonstrated substantially enhanced stimulated

Raman scattering<sup>5,6</sup>, and which exhibit high performance, excellent long-term pressure stability and ease of use. To illustrate the practical potential of these structures, we report two different devices: a hydrogen-filled cell for efficient generation of rotational Raman scattering using only quasi-continuous-wave laser pulses; and acetylene-filled cells, which we use for absolute frequency-locking of diode lasers with very high signal-to-noise ratios. The stable performance of these compact gas-phase devices could permit, for example, gas-phase laser devices incorporated in a ‘credit card’ or even in a laser pointer.

Our all-fibre gas cells consist of hollow-core photonic crystal fibre (HC-PCF) filled with gas and spliced hermetically at both ends to standard single-mode optical fibre (SMF) (Fig. 1a, d). These novel devices have no bulk-optics components, and may be used with a wide range of commercially available optical fibre components (such as couplers, filters, mirrors and lasers). Moreover, the use of HC-PCF in which light is guided in a single transverse mode by means of a photonic bandgap created in the ‘photonic crystal’ cladding<sup>7</sup> greatly increases the efficiency of these laser–gas devices<sup>5,6</sup>.

The high air-filling fraction of the HC-PCFs (which are formed by a network of glass webs typically just a few hundred nanometres thick) makes it very challenging to fusion-splice these fibres without collapsing and deforming the microstructure, which would give a high insertion loss. It is also important to avoid contaminating the PCF with solid deposits and water. Nevertheless, after some practice, routine splicing of HC-PCF to SMF was achieved with a typical loss of 1–2 dB. We estimate that a perfect splicing procedure using our fibres would yield a loss of 0.6–0.8 dB. Of this, 0.15 dB arises from the refractive index mismatch between the fibre cores, and is therefore fundamental. The rest is due to modal field mismatch, which we estimate by the butt-joint approximation<sup>8</sup> to be 0.4–0.6 dB. The discrepancy between estimated and achieved splice losses is linked to the formation of a recess in the end face of the HC-PCF when heated in the splicer (Fig. 1b). We believe this results from the action of surface tension along the many glass–air interfaces within the holey structure, the viscosity of which offers much less resistance to deformation than the solid glass in the outer cladding of the fibre. The mechanical strength of the splices is



**Figure 1** HC-PCF-based gas-cell assembly. **a–c**, Images obtained using a scanning electron microscope. **a**, Side view of a 1,550 nm HC-PCF (the narrower fibre) spliced to an SMF. **b**, End view of an HC-PCF cleaved at the junction of the splice. The recess, which creates an air gap of a few tens of micrometres between the fibre cores, is due to the action of surface tension during fusion. **c**, View of the same piece of HC-PCF as in **a** and **b** but cleaved a few millimetres from the splice, showing clearly the preservation of the microstructural integrity. **d**, Photograph of a 5-m-long hydrogen-filled HC-PCF gas cell, showing its size compared to that of a match.

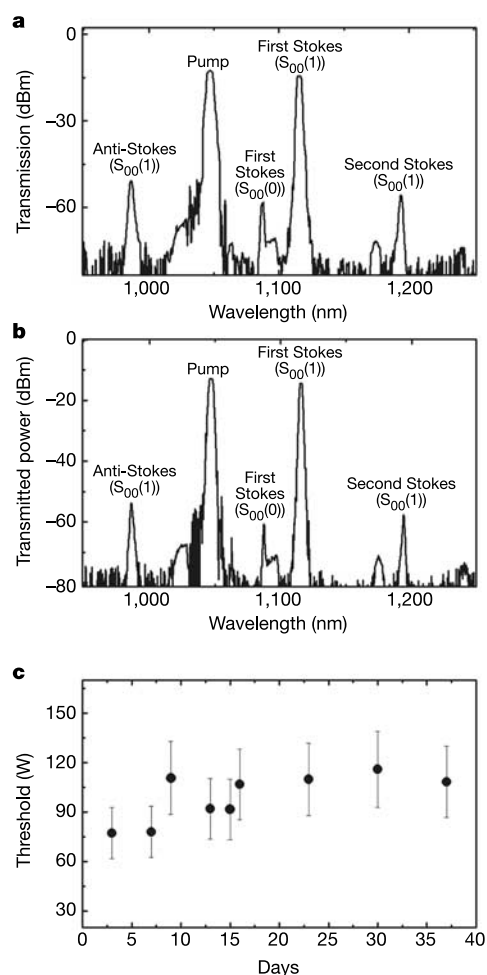
equivalent to a gas pressure of 80 bar—of crucial importance for containment of gases at high pressure.

To form a gas cell, one end of the HC-PCF was fusion-spliced to SMF (see Methods). The HC-PCF was then evacuated for several hours (depending on the length) from the remaining open end to a pressure of  $\sim 1 \mu\text{bar}$ , before being backfilled with the desired gas as described in our previous work<sup>5</sup>. Finally, once the desired fill pressure was achieved, the second end of the HC-PCF was spliced to SMF in the same way as the first splice. Using this procedure, several hydrogen cells and several acetylene cells (Fig. 1d) were produced, with gas fill pressures (before splicing) ranging from 6 mbar to 500 mbar for the former, and from 5 bar to 10 bar for the latter. The acetylene cells were made from HC-PCF with a guidance band centred at 1,550 nm and a loss of  $18 \text{ dB km}^{-1}$ . The hydrogen cells were made from the same material with a guidance band centred at 1,064 nm and a transmission loss of  $60\text{--}70 \text{ dB km}^{-1}$ . In all

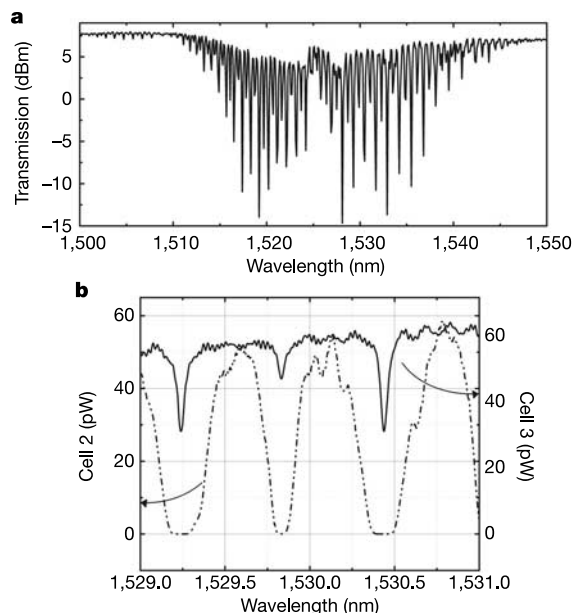
cases, the splice losses were 1–2 dB. Although both gases were highly flammable, no damage to the fibre was apparent during the second splice.

In stimulated Raman scattering (SRS), photons of the incident ‘pump’ laser beam are converted to photons of lower frequency through interactions with molecules in an excited (vibrational or rotational) state. This conversion process becomes highly efficient in HC-PCF, where high intensities and long interaction lengths are available. This has removed the need for powerful lasers, making SRS an ideal technique for efficient laser-frequency conversion and high-resolution spectroscopy. Here we use an experimental set-up (similar to one described previously<sup>5,6</sup>) to generate rotational SRS in a 5-m-long in-line hydrogen gas cell with quasi-continuous-wave pulses. The pump source is a Q-switched Nd:YVO<sub>4</sub> laser at 1,047 nm, with a pulse width tunable in the range 6–50 ns. In all the hydrogen cells we made, we were able to generate SRS with pump pulses as long as 35 ns at peak powers of only 100–180 W. Figure 2a and b shows the typical measured spectra at day 1 and day 37, respectively, when a pulse of 13 ns duration and peak power  $\sim 200 \text{ W}$  is launched into a hydrogen cell with a fill pressure of  $\sim 6 \text{ bar}$  and a splice loss of 1.6 dB and 2 dB for the input and output splices, respectively. The richness of the generated spectrum at such low peak powers illustrates the extreme effectiveness of the gas cell as a Raman converter.

The lifetime of such a cell has been assessed by carrying out threshold measurements on a daily basis on our hydrogen cells (see Methods). Figure 2c shows the evolution with time of the measured threshold peak power (7-ns pulses). The average measured threshold was  $98 \pm 14 \text{ W}$  over a period exceeding 1 month. The measured spectra are identical within the uncertainty of our measurement, indicating good stability of the gas cell over the measurement period. In principle, with a strong seal, the hydrogen will escape only via diffusion through the silica, which in our case (isothermal conditions at room temperature and for a fill pressure of  $\sim 6 \text{ bar}$ ) would yield a leakage rate of less than  $0.001 \text{ mol yr}^{-1}$ .



**Figure 2** The SRS spectrum generated by a hydrogen-filled HC-PCF gas cell, and its behaviour as a function of time. **a**, **b**, Rotational SRS spectra generated using a 13-ns pulse with a peak power in the range of 150–200 W at day 1 (**a**) and day 37 (**b**). For both traces, in addition to the pump line, the spectrum contains three lines from the rotational transition  $S_{00}(1)$  in ortho-hydrogen; the first Stokes ( $\sim 1,115 \text{ nm}$ ), the second Stokes ( $\sim 1,194 \text{ nm}$ ), and finally the first anti-Stokes at  $\sim 986 \text{ nm}$ , which is the product of wave mixing between the pump and the first Stokes. Also, the spectrum exhibits a peak at  $\sim 1,087 \text{ nm}$ , which corresponds to the first Stokes line of the much weaker Raman transition,  $S_{00}(0)$ , from the para-hydrogen present ( $\sim 30\%$ ) in normal hydrogen gas. **c**, The variation with time of the peak power of the pump threshold for the generation of the first Stokes of  $S_{00}(1)$ . The error bars indicate the standard error on the threshold measurement.



**Figure 3** Transmission spectra of acetylene-filled HC-PCF gas cells. **a**, Transmission spectrum through a 1-m-long HC-PCF-based  $^{12}\text{C}_2\text{H}_2$  cell. The spectrum exhibits more than 50 strong lines, thereby providing a wide grid of frequency references. **b**, Close-up view of the transmission spectrum of cell 2 (fill pressure  $\sim 200 \text{ mbar}$ ) and cell 3 (fill pressure  $\sim 6 \text{ mbar}$ ) with a resolution of  $0.01 \text{ nm}$  around  $1,530 \text{ nm}$ . The linewidths in cell 3 are Doppler limited.



(ref. 9), suggesting that the lifetime of the cell could extend to several decades.

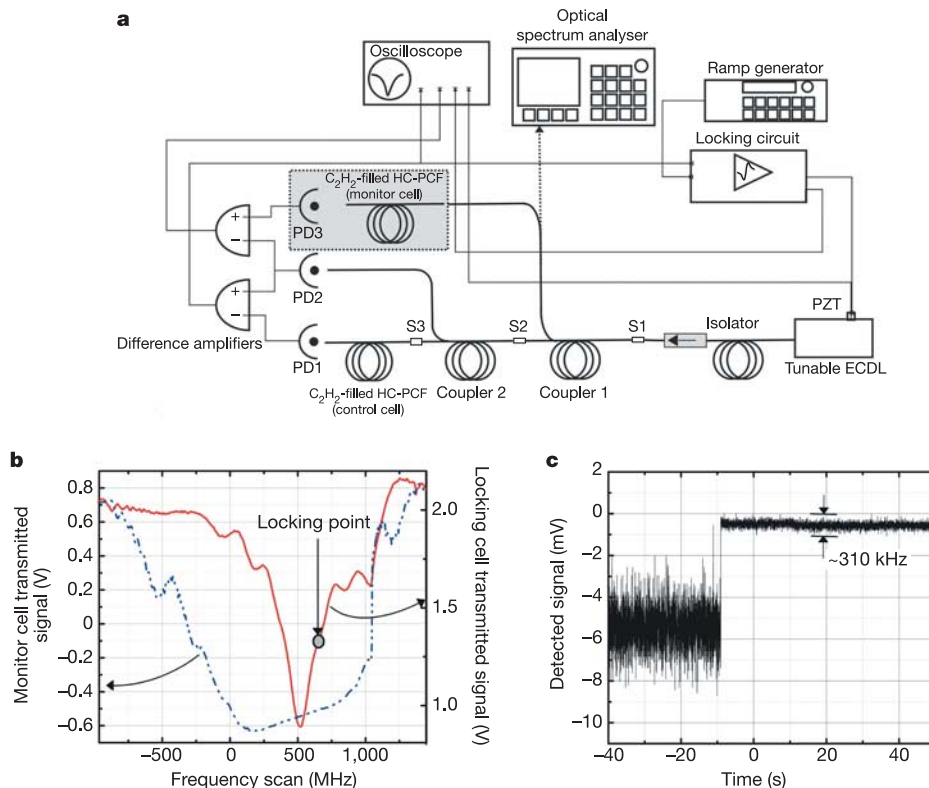
Accurate stabilization of laser frequency is essential in many fields<sup>1,2</sup>, such as high-resolution spectroscopy, measurements of fundamental physical constants<sup>10,11</sup>, atomic physics and optical telecommunications; in all these cases, there is a pressing need for wavelength accuracy and stability to reduce the spacing between the channels, and increase their number, in wavelength division multiplexed systems. This can be achieved by locking the laser frequency to an optical frequency reference—usually an ensemble of atoms or molecules with a narrow absorption line. The performance of the stabilization system is limited by the fractional frequency fluctuations. One needs a high quality factor,  $Q$  ( $Q = \nu/\Delta\nu$ , where  $\nu$  is the carrier frequency and  $\Delta\nu$  is the linewidth of the reference line), and a high signal-to-noise ratio, which is in turn determined by the stabilization detection scheme and the interaction length between the light and the absorber. The long interaction lengths offered by HC-PCF<sup>5,6</sup> permit large enhancements in signal-to-noise ratio.

Acetylene is suitable for the optical communications wavelength range<sup>12</sup>, offering a comb of stable and regularly spaced overtone absorptions in the vicinity of 1.55  $\mu\text{m}$ . It exhibits Doppler-free saturation linewidths of less than 1 MHz (ref. 13) and is remarkably insensitive to external perturbations<sup>14</sup>. Figure 3a shows the typical transmitted spectrum of a broad LED (light-emitting-diode) light source through a 1-m-long HC-PCF-based acetylene cell. Figure 3b shows a more detailed spectrum around the 1,530.5-nm line transmitted through two cells (labelled cell 3 and cell 2) of the same length (1 m) but containing acetylene at different pressures

( $\sim 6$  mbar for cell 2 and  $\sim 200$  mbar for cell 3). Most of the strong lines exhibit more than 90% absorbance.

For illustrative purposes, an all-fibre frequency stabilization system was built and tested (Fig. 4a). A simple side-locking technique was used to lock to one of the shoulders of a selected absorption line<sup>15</sup>. This approach, though it does not offer the best performance, has the merit of being simple and modulation-free. The set-up consisted of a tunable extended cavity diode laser (ECDL) source, an isolator, two couplers and a length of acetylene-filled HC-PCF. The laser output, after passing the isolator, is split into three parts using two fused taper couplers. The first (the locking beam) passes through the gas cell (cell 3) and is detected. The second (the reference beam) is monitored at an identical detector. The reference and locking signals are sent to a difference amplifier before being fed to a locking circuit based on ref. 15. Before switching on the control loop, the wavelength of the laser is tuned to the desired absorption line. To avoid possible artefacts in the control system<sup>16</sup>, it is important to have an independent means of testing (discriminating) the absolute frequency. This was achieved by passing the third (out-of-loop) part of the laser beam through a second HC-PCF acetylene cell ('monitor cell' in Fig. 4a). This resulted in a frequency stabilization and testing system that was completely fibre-based—to our knowledge, the first time this has been shown. Using this system, we were able to lock the laser to different acetylene absorption lines.

This is illustrated in Fig. 4b and c. Figure 4b shows the detected transmission at 1,530.43 nm (P9 absorption<sup>14</sup>) from the locking and monitor cells as the laser frequency is swept through 2 GHz. Figure 4c shows the frequency fluctuations of the laser when it is free



**Figure 4** All-fibre frequency stabilization and discrimination system using acetylene ( $\text{C}_2\text{H}_2$ )-filled cells. **a**, Schematic representation of the frequency-locking set-up. PD, photodetector; S, splice; ECDL, extended cavity diode laser. PZT, piezo-electrical transducer. **b**, The red solid curve shows the detected absorption peaks (P9 line) after transmission through the control cell, and the blue dot-dash curve those detected after the monitor cell. The asymmetric shape of the P9 absorption peak measured at the monitor

cell is caused by weak excited state absorption lines<sup>17</sup> and the etalon effect of the gas cell, and is consistent with the trace from the optical spectrum analyser. The locking point represents the frequency at which the laser is locked. **c**, The frequency fluctuations of the laser when free running (left-hand part of the trace) and when locked, as monitored by the out-of-loop acetylene cell. The frequency fluctuations of the laser when the locking loop is closed exhibit a maximum r.m.s. frequency deviation of  $\sim 310$  kHz.

running and when the locking loop is closed. The closed-loop frequency fluctuations exhibit a root-mean-square (r.m.s.) frequency deviation of  $\sim 310$  kHz over a 1-min integration time. The residual noise exhibits a white noise spectrum with an Allan variance of  $3 \times 10^{-11}/\tau^{1/2}$  for an averaging time of  $1 \text{ s} \leq \tau \leq 25.6 \text{ s}$ , and is limited by the electronic noise of our detection system (PD3 in Fig. 4a).

A straightforward improvement would be to use a saturable absorption line (which can have a linewidth as narrow as 1 MHz; ref. 13) as a reference. The improved signal-to-noise ratio in HC-PCF also makes overtone absorptions in the visible and near-infrared (for example, P11 of  $^{12}\text{C}_2\text{H}_2$  at 790.703 nm) accessible to laser frequency metrology. To test for gas leakage, the pressure-dependent linewidth of the P9 acetylene absorption was monitored daily over two months. The results gave an average value of  $\sim 468$  MHz with a standard deviation of 30 MHz, that is, the linewidth is consistent with the Doppler limited value and was constant within experimental error, indicating no measurable leakage of gas.

In conclusion, all-fibre, ultra-compact, high performance, easy-to-use and unconditionally stable gas-laser devices have been reported. The commercial availability of a wide range of all-fibre components (for example, lasers, phase modulators, power attenuators, isolators, Bragg gratings and beam-splitters) makes complex systems easy to design and construct. It is now possible to imagine miniature laser-gas devices, occupying a tiny volume and containing minute amounts of gas, with everyday applications in fields such as the colour conversion of laser light (perhaps using a built-in diode pump laser), and the measurement and stabilization of laser frequency. The unique features of HC-PCFs make these gas-laser devices extremely efficient, and their impact in many laser-related fields is likely to be deep and lasting. □

## Methods

### Splicing procedure

Fusion-splicing was carried out using a commercial filament-based splicer (Vytran FFS-2000-PM). In this splicer, the splicing region is continuously purged by argon gas to stop the filament burning. This prevents contamination of the splice by solid deposits and water condensation, and also prevents combustion of flammable gases such as hydrogen and acetylene. For high fill pressures, the splice could be consolidated using heat-curable glue.

### Threshold measurements

The energy threshold measurement for the generation of the first Stokes line  $S_{00}(1)$  was carried out in the following manner. The input power was varied by manual rotation of a half-wave plate, and the output light was collected and fed to an optical spectrum analyser and power detectors. The detected signal at the Stokes wavelength was monitored as the input power was increased up to a point where the Stokes signal emerged abruptly from the noise on the optical spectrum analyser. With this technique, we found a relative uncertainty varying between 40% and 70%.

Received 3 November 2004; accepted 6 January 2005; doi:10.1038/nature03349.

1. Luiten, A. N. (ed.) *Frequency Measurement and Control* (Springer, Berlin, 2001).
2. Hall, J. L. Frequency stabilized lasers—a parochial review. *Proc. SPIE* **1837**, 2–15 (1992).
3. Hentschel, M. et al. Attosecond metrology. *Nature* **414**, 509–513 (2001).
4. Steinmeyer, G., Sutter, D. H., Gallmann, L., Matuschek, N. & Keller, U. Frontiers in ultrashort pulse generation: Pushing the limits in linear and nonlinear optics. *Science* **286**, 1507–1512 (1999).
5. Benabid, F., Knight, J. C., Antonopoulos, G. & Russell, P. S. J. Stimulated Raman scattering in hydrogen-filled hollow-core photonic crystal fiber. *Science* **298**, 399–402 (2002).
6. Benabid, F., Bouwmans, G., Knight, J. C., Russell, P. S. J. & Couy, F. Ultra-high efficiency laser wavelength conversion in gas-filled hollow core photonic crystal fiber by pure stimulated rotational Raman scattering in molecular hydrogen. *Phys. Rev. Lett.* **93**, 123903 (2004).
7. Russell, P. Photonic crystal fibers. *Science* **299**, 358–362 (2003).
8. Marcuse, D. Loss analysis of single-mode fiber splices. *Bell Syst. Tech. J.* **55**, 703–718 (1977).
9. Dushman, S. *Scientific Foundations of Vacuum Technique* (John Wiley & Sons, New York, 1962).
10. Hansch, T. W. in *The Hydrogen Atom* (ed. Hansch, T. W.) 93–102 (Springer, Berlin, 1989).
11. Huber, A. et al. Hydrogen-deuterium 1S–2S isotope shift and the structure of the deuteron. *Phys. Rev. Lett.* **80**, 468–471 (1998).
12. Quinn, T. J. Practical realization of the definition of the metre, including recommended radiations of other optical frequency standards (2001). *Metrologia* **40**, 103–133 (2003).
13. Nakagawa, K., de Labachellerie, M., Awaji, Y. & Kourogi, M. Accurate optical frequency atlas of the 1.5  $\mu\text{m}$  bands of acetylene. *J. Opt. Soc. Am. B* **13**, 2708–2714 (1996).
14. Swann, W. C. & Gilbert, S. L. Pressure-induced shift and broadening of 1510–1540-nm acetylene wavelength calibration lines. *J. Opt. Soc. Am. B* **17**, 1263–1270 (2000).

15. MacAdam, K. B., Steinbach, A. & Wieman, C. A narrow band tunable diode laser system with grating feedback, and a saturated absorption spectrometer for Cs and Rb. *Am. J. Phys.* **69**, 1098–1111 (1992).
16. Wieman, C. E. & Hollberg, L. Using diode lasers for atomic physics. *Rev. Sci. Instrum.* **62**, 1–20 (1991).
17. Onae, A. et al. Optical frequency link between an acetylene stabilized laser at 1542 nm and an Rb stabilized laser at 778 nm using a two-color mode-locked fiber laser. *Opt. Commun.* **183**, 181–187 (2000).

**Acknowledgements** The support of the UK Engineering and Physical Sciences Research Council (EPSRC) is acknowledged. F.B. thanks A. Luiten and M. Maric for discussions on frequency stabilization. F.B. is an EPSRC Advanced Research Fellow.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to F.B. (pysab@bath.ac.uk).

## Obliquity pacing of the late Pleistocene glacial terminations

Peter Huybers<sup>1</sup> & Carl Wunsch<sup>2</sup>

<sup>1</sup>Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

<sup>2</sup>Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The 100,000-year timescale in the glacial/interglacial cycles of the late Pleistocene epoch (the past  $\sim 700,000$  years) is commonly attributed to control by variations in the Earth's orbit<sup>1</sup>. This hypothesis has inspired models that depend on the Earth's obliquity ( $\sim 40,000$  yr;  $\sim 40$  kyr), orbital eccentricity ( $\sim 100$  kyr) and precessional ( $\sim 20$  kyr) fluctuations<sup>2–5</sup>, with the emphasis usually on eccentricity and precessional forcing. According to a contrasting hypothesis, the glacial cycles arise primarily because of random internal climate variability<sup>6–8</sup>. Taking these two perspectives together, there are currently more than thirty different models of the seven late-Pleistocene glacial cycles<sup>9</sup>. Here we present a statistical test of the orbital forcing hypothesis, focusing on the rapid deglaciation events known as terminations<sup>10,11</sup>. According to our analysis, the null hypothesis that glacial terminations are independent of obliquity can be rejected at the 5% significance level, whereas the corresponding null hypotheses for eccentricity and precession cannot be rejected. The simplest inference consistent with the test results is that the ice sheets terminated every second or third obliquity cycle at times of high obliquity, similar to the original proposal by Milankovitch<sup>12</sup>. We also present simple stochastic and deterministic models that describe the timing of the late-Pleistocene glacial terminations purely in terms of obliquity forcing.

To test whether the glacial variability is related to changes in Earth's astronomical configuration, we adopt a formal null-hypothesis ( $H_0$ ) that glacial terminations are independent of obliquity variations, and the alternative hypothesis ( $H_1$ ) that glacial terminations are paced by it. Our focus on obliquity is motivated by previous indications of nonlinear interactions between obliquity period and quasi-100-kyr glacial variability<sup>13</sup>, but we also perform identical tests for pacing by precession and eccentricity. The test is focused on glacial terminations because their magnitude and abruptness facilitate accurate identification.

Several obstacles must be overcome to distinguish between  $H_0$  and  $H_1$ . A major problem is the need to establish time controls on the glacial variability. Many studies estimate age by assuming a relationship between climate proxy variability and orbital

forcing<sup>14,15</sup>, but this approach assumes the validity of the hypothesis being tested. Instead, we use an age-model devoid of orbital assumptions and apply it to the leading empirical orthogonal function (EOF1) of ten well-resolved marine  $\delta^{18}\text{O}$  records<sup>13</sup>, a proxy for ice volume and ocean temperature (Fig. 1a).

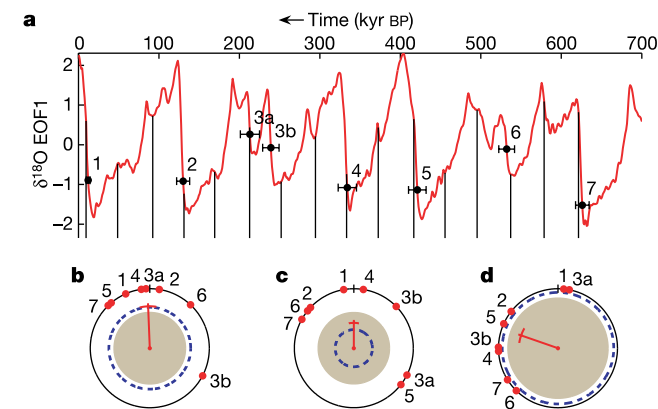
Most simple models of the late Pleistocene glacial cycles have at least four degrees of freedom<sup>2</sup>, and some have as many as twelve<sup>3</sup>. Unsurprisingly then, the seven observed quasi-100-kyr glacial cycles are insufficient to distinguish between the skill of the various models<sup>16</sup>. Models with minimal degrees of freedom are necessary. Other requirements for a useful test include the ability to cope with noisy records, age-model uncertainty and (possibly) nonlinear interactions. Here we test for stability in the phase of the orbital parameters during glacial terminations using Rayleigh's  $R$  (see Methods).

To proceed, we must estimate the probability distribution function associated with  $H_0$ . Of the five estimation methods explored, the one adopted gives the highest critical value and thus makes  $H_0$  the most difficult to reject—a modified random walk<sup>8</sup> representing ice-volume variability:

$$V_{t+1} = V_t + \eta_t \text{ and if } V_t \geq T_o, \text{ terminate.} \quad (1)$$

This highly simplified model posits 1-kyr steps in ice volume,  $V_t$ , of random length,  $\eta_t$ , independently drawn from a normal distribution with standard deviation  $\sigma = 2$  and mean  $\mu = 1$ . The non-zero mean biases the Earth toward glaciation. Once  $V_t$  reaches a threshold,  $T_o = 90$ , a termination is triggered, and ice-volume is linearly reset to zero over 10 kyr. If the model were deterministic with  $\sigma = 0$  and  $\mu = 1$ , glacial cycles would last exactly 100 kyr, but with  $\sigma = 2$ , glacial cycle duration is approximately normally distributed at  $(100 \pm 20)$  kyr, a spread consistent with observations<sup>11</sup>. Initial ice volume is randomly set between 0 and  $T_o$  with equal probability. From a Monte Carlo technique (see Methods), we find a critical value of  $R = 0.60$ .

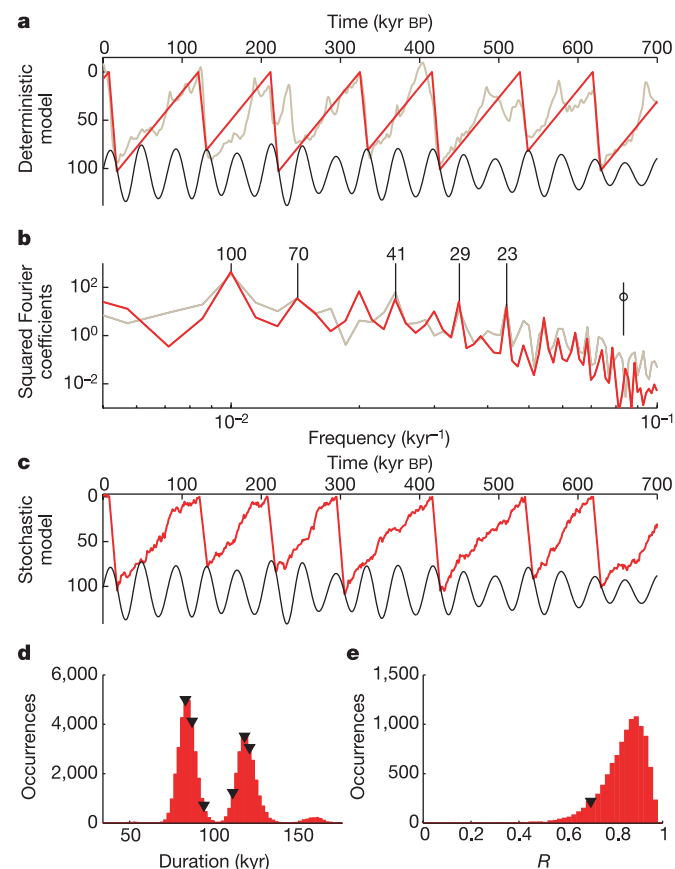
The observed obliquity phases produce  $R = 0.70$ , and  $H_0$  is rejected (Fig. 1b). This rejection of  $H_0$  is robust to all plausible



**Figure 1** The Rayleigh test for phase directionality. **a**,  $\delta^{18}\text{O}$  EOF1 normalized so that negative values indicate more ice. Dots indicate onset of a termination and horizontal bars indicate one-standard-deviation age-model uncertainties<sup>13</sup>. Termination 3 is split between events 3a and 3b. Vertical lines indicate the time of maxima in obliquity. Note that time runs from right to left in the palaeoclimate convention. **b**, The obliquity phase (dots) sampled at each termination and plotted on a unit circle. The vector average has a magnitude  $R = 0.70$  (cross mark) exceeding the critical value  $c = 0.60$  (filled circle), so that  $H_0$  is rejected. Furthermore,  $R$  is near  $H_1$ 's maximum-likelihood value (dashed circle). The direction is indistinguishable from maximum obliquity (top of the circle). **c, d**, Analogous tests are made for precession ( $R = 0.43$ ,  $c = 0.60$ ) (**c**) and eccentricity ( $R = 0.66$ ,  $c = 0.84$ ) (**d**), but in neither case can the corresponding  $H_0$  be rejected. See the Supplementary Information for more details.

reformulations of the test. Thus, the phase of obliquity has a statistically significant relationship with the timing of deglaciation. The mean phase at deglaciation is indistinguishable from zero and is associated with maxima in obliquity. We estimate the  $H_1$  probability density function by assuming that terminations always initiate at the same phase of obliquity, but that termination timing is subject to identification and age-model uncertainties (see Methods). The maximum-likelihood value of  $H_1$  is  $R = 0.69$ , very near the observed value, further supporting the conclusion that glacial terminations are paced by variations in obliquity. Hypotheses not accounting for the obliquity pacing are unlikely to be correct.

Analogous hypothesis tests for precession and eccentricity produce different results (Fig. 1c, d). Age-model uncertainty approaches half a precession cycle, so that the power of the precession test is negligible—even if present, precession pacing of the glacial cycles cannot be discerned. In the case of eccentricity,  $H_0$  is not rejected using the random-walk probability estimate (equation (1)), but is rejected using weaker formulations of the eccentricity null hypothesis. The discrepancy arises because null



**Figure 2** Deterministic and stochastic descriptions of the late-Pleistocene glacial variability. **a**, Deterministic model results (red) with an obliquity-dependent threshold (black) plotted over EOF1 (brown). **b**, Periodograms of the deterministic model results (red) and EOF1 (brown). Concentrations of energy are centred on the 1/41-kyr obliquity frequency and the 1/100-kyr glacial band; as well as combination tones at 1/70, 1/29 and 1/23 kyr. The approximate 95% confidence interval is indicated by the vertical bar on the right. **c**, A realization of the stochastic model. **d**, Histogram of the time between terminations, derived from many runs of the stochastic model. The observed duration between terminations (triangles, using termination 3a not 3b) coincide with the dominant 80- and 120-kyr modes. **e**, Histogram of Rayleigh's  $R$  from the stochastic model with the observed obliquity value,  $R = 0.70$ , indicated by the triangle.



hypotheses that assume a glacial timescale of roughly 100 kyr (which seem more physical) tend to have higher  $R$  values and are more difficult to reject. Because the hypotheses of negligible influence of precession and eccentricity on the glacial terminations cannot be rejected, we adopt a minimalist strategy, retaining only obliquity to describe the glacial terminations.

From a physical standpoint, support for the obliquity control hypothesis also comes from the fact that maxima in obliquity cause annual average insolation anomalies of up to  $10 \text{ W m}^{-2}$  at high latitudes. Furthermore, the annual average and seasonal insolation redistributions associated with obliquity are hemispherically symmetric—as is the glacial variability to within a few thousand years<sup>17,18</sup>. Finally, although obliquity control does not by itself address the question of why termination five is anomalously large, it does alleviate the marine isotope stage 11 problem<sup>2</sup> of explaining the initiation of termination five when eccentricity and precession amplitude are small.

But how does a forcing with a 40-kyr period pace the 100-kyr late-Pleistocene glacial variability? One suggestion is that the phase and amplitude modulations of obliquity cause the 100-kyr variability<sup>4</sup>, but it is difficult to see the climatic significance of such modulations. Instead, we suggest that the climate state skips one or two obliquity beats before deglaciating, thus giving quantized glacial-cycle durations of either 80 or 120 kyr. A speculative scenario is for increased obliquity to increase high-latitude insolation and cause heating of an ice sheet, eventually warming the ice–bedrock interface. When the ice sheet is thin, basal temperature and pressure are low<sup>19</sup>, and the obliquity heating has little effect—a skipped beat. But when the ice sheet is thick, basal temperature and pressure are high<sup>19</sup>, and additional obliquity heating promotes basal melting. Enhanced lubrication of the ice–bedrock interface may trigger deglaciation by increasing ice flux into the ocean or toward lower latitudes, as well as by increasing the thinning rate and causing inward migration of the ablation zone<sup>20</sup>. Note that  $\sim 10$  kyr are required for surface heating to penetrate to the base of an ice sheet<sup>19</sup>. Unlike precession, changes in obliquity are associated with sustained annual average changes in insolation<sup>21,22</sup> and are thus more likely to cause basal warming. Furthermore, if climate was warmer during the early Pleistocene, terminations would be triggered more nearly every obliquity cycle, giving the observed smaller and more rapid glacial variability<sup>23</sup>.

A simple system, consistent with the above observations, is obtained by making the threshold in equation (2) dependent on obliquity:

$$V_{t+1} = V_t + \eta_t \text{ and if } V_t \geq T_o - a\bar{\theta}_t, \text{ terminate.} \quad (2)$$

Here  $\bar{\theta}_t$  is obliquity<sup>24</sup>, normalized to zero mean and unit variance, with amplitude  $a$ . The time-variable threshold makes it more likely for a termination to occur when obliquity is large. Both deterministic and stochastic variants of equation (2) are offered, to emphasize that such models are not theories of climate change, but rather attempts at efficient kinematic descriptions of the data, and that different mechanisms can be consistent with the limited observational records.

To make the model deterministic, the ice-volume step size in equation (2) is fixed at  $\eta = 1$ . Setting  $a = 15$ ,  $T_o = 105$  and (initial ice volume)  $V_{(t=-700)} = 30$  provides a good description of the late-Pleistocene glacial variability (Fig. 2a). Many other parameterizations yield similar results—we choose this one because it is particularly simple. The model produces the correct timing of the glacial terminations (using termination 3a, not 3b) and has a squared-cross-correlation with  $\delta^{18}\text{O}$  EOF1 of 0.65, an excellent fit considering there are only three adjustable parameters. (Parameters other than  $\eta$  could be adjusted instead.) For comparison, tuning other models having four<sup>2</sup> or twelve<sup>3</sup> adjustable parameters yields a maximum squared-cross-correlation with EOF1 of 0.24 and 0.74

respectively<sup>25</sup>. If skill is measured by squared-cross-correlation divided by the number of adjustable parameters, equation (2) does more than three times as well.

A periodogram of the deterministic model results (Fig. 2b) shows narrow-band concentrations of energy at the average 100-kyr period, the 41-kyr obliquity period, and at previously identified<sup>13</sup> combination tones  $1/41 - 1/100 = 1/70$ ,  $1/41 + 1/100 = 1/29$ , and  $1/41 + 2/100 = 1/23$  kyr—in good agreement with the  $\delta^{18}\text{O}$  EOF1 periodogram. The appearance of 1/23-kyr narrow-band energy, in the absence of precession-band forcing, highlights the ambiguous origins of this energy band<sup>22</sup>. The model also has energy at 2/100 kyr, not visible in the observations. Note that the model produces an energetic background continuum consistent with the  $\delta^{18}\text{O}$  periodogram even though it is deterministic.

For the stochastic case,  $\eta_t$  is defined to be normally distributed with  $\sigma = 2$  and  $\mu = 1$ . Here  $\eta_t$  represents the unpredictable background weather and climate variability spanning all timescales out to the glacial/interglacial. All other parameter settings are kept the same as in the deterministic case. Now, equation (2) resembles an order-one autoregressive process. Such a process is known to describe the glacial variability well, except for runs associated with glacial terminations<sup>26</sup>, modelled here using the threshold condition in equation (2). The time between terminations in the stochastic model averages 100 kyr but has a tri-modal distribution with maxima at two (80 kyr), three (120 kyr), and four (160 kyr) obliquity cycles (Fig. 2d). The observed durations between terminations are consistent with the dominant 80- and 120-kyr modes, but suggest that one would see 160-kyr glacial cycles, given a larger sample size. The  $R$  value obtained by the stochastic model averages 0.85 (Fig. 2e), higher than the observed  $R = 0.70$ ; this is as expected because of observational age-model error.

Deterministic and stochastic variants of equation (2) thus both describe the late-Pleistocene glacial variability. A description of the early Pleistocene variability<sup>23</sup> is obtained by lowering the termination threshold to  $T_o = 40$ , giving smaller-amplitude terminations that occur more nearly every obliquity cycle. Alternatively, instead of specifying a parameter change, the mid-Pleistocene transition can be described using a chaotic model<sup>25</sup> (not shown) having spontaneous transitions between 40 and 100-kyr modes of glacial variability. With only seven cycles, it is unclear whether adequate data will ever be available to distinguish between stochastic, simple deterministic and chaotic deterministic models of the glacial variability.

The simplest interpretation of our results is that during the Pleistocene epoch, the Earth tends to a glacial state (anthropogenic influences aside) and deglaciates near some, but not all, obliquity maxima. Obliquity control of the timing of deglaciation, probably in concert with a stochastic forcing, has several other consequences. These include inferences drawn from precession-based and eccentricity-based models of glaciation<sup>27</sup>, the hemispheric symmetry of glacial cycles, and the efficacy of age-model tuning of cores. These issues will be taken up elsewhere.  $\square$

## Methods

Rayleigh's  $R$  is defined as<sup>28</sup>:

$$R = \frac{1}{N} \left| \sum_{n=1}^N \cos \phi_n + i \sin \phi_n \right| \quad (3)$$

Here,  $\phi_n$  is the phase of obliquity stroboscopically sampled at the  $n$ th glacial termination. The line brackets indicate the magnitude, making  $R$  real and non-negative, with a maximum value of one when the phases are all equal. Relative to measures of phase coupling used to investigate other nonlinear interactions<sup>29</sup>, Rayleigh's  $R$  requires many fewer phases (roughly five) to test for phase stability<sup>28</sup>.

To measure  $R$  between obliquity and the glacial cycles, we use  $\delta^{18}\text{O}$  EOF1 and an age-model independent of orbital assumptions<sup>13</sup> (Fig. 1a). Such an independent age-model is important because even so-called minimal tuning strategies—using only a narrow band of a climate record—tend to align the abrupt glacial terminations with a particular phase of the assumed forcing (as indicated by Monte Carlo simulations), thus biasing records towards  $H_1$ .

The rate of change of EOF1 has a unimodal distribution with a long tail, indicative of rapid melting. Terminations are defined to initiate when the rate of change in EOF1 first exceeds the two-standard-deviation level. This criterion identifies each of the usual terminations<sup>10,11</sup>, and two events in the termination 3 deglacial sequence (termed 3a and 3b for the younger and older events, respectively). Additional rules could be added to exclude 3a or 3b, but this rejection seems ad hoc, and so we use all eight termination events. Note, the timing of termination 3 predicted by the Paillard model<sup>3</sup> also coincides with either event 3a or 3b, depending on slight changes in the parameterizations. Reassuringly, results are not sensitive to details of the test: either termination 3a or 3b can be excluded; termination times can be defined using the midpoint between the local minimum and maximum bracketing each termination; and individual benthic or planktic  $\delta^{18}\text{O}$  records<sup>15</sup> can be used in place of EOF1.

The probability density function (PDF) associated with  $H_0$  is estimated using the modified random-walk model (equation (1)). A realization of  $R$  is obtained by sampling the phase of obliquity at eight consecutive termination initiations, generated from equation (1), and the PDF of  $H_0$  is estimated by binning  $10^4$  such realizations of  $R$ . Other methods are to assume a uniform phase distribution, use surrogate data techniques<sup>30</sup>, or to derive statistics from ensemble runs of other models, but all of these give PDF estimates which make  $H_0$  more easily rejected and are therefore not used.

To estimate the PDF associated with  $H_1$ , we assume that glacial terminations always occur at the same phase of obliquity, but that the phase observations are subject to identification and age-model error. A Monte Carlo technique is used where the timing of the glacial terminations are perturbed according to the estimated age uncertainties<sup>13</sup> (these average  $\pm 9$  kyr) and identification error ( $\pm 1$  kyr, the EOF1 sampling resolution). A realization of  $R$  is then computed using the phase of obliquity at the perturbed ages, and  $10^4$  such realizations are binned to estimate the PDF of  $H_1$ . We estimate the likelihood of correctly rejecting  $H_0$  (that is, the power of the obliquity test) to be 0.57. See the Supplementary Information for a listing of the other pertinent data and statistics.

Received 6 August 2004; accepted 24 January 2005; doi:10.1038/nature03401.

- Hays, J., Imbrie, J. & Shackleton, N. Variations in the earth's orbit: Pacemaker of the ice ages. *Science* **194**, 1121–1132 (1976).
- Imbrie, J. & Imbrie, J. Modeling the climatic response to orbital variations. *Science* **207**, 943–953 (1980).
- Paillard, D. The timing of Pleistocene glaciations from a simple multiple-state climate model. *Nature* **391**, 378–391 (1998).
- Liu, H. Insolation changes caused by combination of amplitude and frequency modulation of the obliquity. *J. Geophys. Res.* **104**, 25197–25206 (1999).
- Gildor, H. & Tziperman, E. Sea ice as the glacial cycles' climate switch: Role of seasonal and orbital forcing. *Paleoceanography* **15**, 605–615 (2000).
- Saltzman, B. Stochastically-driven climatic fluctuations in the sea-ice, ocean temperature,  $\text{CO}_2$ , feedback system. *Tellus* **34**, 97–112 (1982).
- Pelletier, J. Coherence resonance and ice ages. *J. Geophys. Res.* **108**, doi:10.1029/2002JD003120 (2003).
- Wunsch, C. The spectral description of climate change including the 100ky energy. *Clim. Dyn.* **20**, 353–363 (2003).
- Saltzman, B. *Dynamical Paleoclimatology: Generalised Theory of Global Climate Change* (Academic, San Diego, 2002).
- Broecker, W. Terminations. in *Milankovitch and Climate* (eds Berger, A. et al.) Part 2, 687–698 (D. Riedel, Hingham, 1984).
- Raymo, M. E. The timing of major climate terminations. *Paleoceanography* **12**, 577–585 (1997).
- Milankovitch, M. *Kanon der Erdbestrahlung und seine Anwendung auf das Eiszeiten-problem* (Royal Serbian Academy, Belgrade, 1941).
- Huybers, P. & Wunsch, C. A depth-derived Pleistocene age-model: Uncertainty estimates, sedimentation variability, and nonlinear climate change. *Paleoceanography* **19**, doi:10.1029/2002PA000857 (2004).
- Imbrie, J. et al. in *Milankovitch and Climate* (eds Berger, A. et al.) Part 1, 269–305 (D. Riedel Publishing Company, 1984).
- Shackleton, N. J., Berger, A. & Peltier, W. R. An alternative astronomical calibration of the lower Pleistocene timescale based on ODP site 677. *Trans. R. Soc. Edinb. Earth Sci.* **81**, 251–261 (1990).
- Roe, G. & Allen, M. A comparison of competing explanations for the 100,000-yr ice age cycle. *Geophys. Res. Lett.* **26**, 2259–2262 (1999).
- Blunier, T. & Brook, E. Timing of millennial-scale climate change in Antarctica and Greenland during the last glacial period. *Science* **291**, 109–112 (2001).
- Wunsch, C. Greenland-Antarctic phase relations and millennial time-scale climate fluctuations in the Greenland cores. *Quat. Sci. Rev.* **22**, 1631–1646 (2003).
- Marshall, S. & Clark, P. Basal temperature evolution of North American ice sheets and implications for the 100-kyr cycle. *Geophys. Res. Lett.* **29**, doi:10.1029/2002GL015192 (2002).
- Zwally, H. et al. Surface melt-induced acceleration of greenland ice-sheet flow. *Science* **297**, 218–222 (2002).
- Rubincam, D. Insolation in terms of earth's orbital parameters. *Theor. Appl. Climatol.* **48**, 195–202 (1994).
- Huybers, P. & Wunsch, C. Rectification and precession-period signals in the climate system. *Geophys. Res. Lett.* **30**, doi:10.1029/2003GL017875 (2003).
- Raymo, M. & Nisancioglu, K. The 41 kyr world: Milankovitch's other unsolved mystery. *Paleoceanography* **18**, doi:10.1029/2002PA000791 (2003).
- Berger, A. & Loutre, M. F. Astronomical solutions for paleoclimate studies over the last 3 million years. *Earth Planet. Sci. Lett.* **111**, 369–382 (1992).
- Huybers, P. On the Origins of the Ice Ages: Insolation Forcing, Age Models, and Nonlinear Climate Change. PhD thesis, MIT (2004).
- Wunsch, C. Quantitative estimate of the Milankovitch-forced contribution to observed quaternary climate change. *Quat. Sci. Rev.* **23**, 1001–1012 (2004).

- Ruddiman, W. F. Orbital insolation, ice volume, and greenhouse gases. *Quat. Sci. Rev.* **22**, 1597–1622 (2003).
- Upton, G. & Fingleton, B. *Spatial Data Analysis by Example Vol. 2* (John Wiley and Sons, Chichester, 1989).
- Rosenblum, M. & Pikovsky, A. Synchronization: from pendulum clocks to chaotic lasers and chemical oscillators. *Contemp. Phys.* **44**, 401–416 (2003).
- Schreiber, T. & Schmitz, A. Surrogate time series. *Physica D* **142**, 346–382 (2000).

Supplementary Information accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Useful comments were provided by E. Boyle, W. Curry, T. Herbert, J. McManus, F. Ng, M. Tingley and G. Yang. P.H. is supported by the NOAA Postdoctoral Program in Climate and Global Change and C.W. is supported in part by the National Ocean Partnership Program (ECCO).

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to P.H. (phuybers@whoi.edu).

## Two episodes of microbial change coupled with Permo/Triassic faunal mass extinction

Shucheng Xie<sup>1,2</sup>, Richard D. Pancost<sup>1</sup>, Hongfu Yin<sup>2</sup>, Hongmei Wang<sup>3</sup> & Richard P. Evershed<sup>1</sup>

<sup>1</sup>Organic Geochemistry Unit, Bristol Biogeochemistry Center, School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK

<sup>2</sup>Faculty of Earth Science and State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences, Wuhan 430074, People's Republic of China

<sup>3</sup>School of Environmental Studies, China University of Geosciences, Wuhan 430074, People's Republic of China

Microbial expansion following faunal mass extinctions in Earth history can be studied by petrographic examination of microbialites (microbial crusts)<sup>1,2</sup> or well-preserved organic-walled microbes<sup>3</sup>. However, where preservation is poor, quantification of microbial communities can be problematic. We have circumvented this problem by adopting a lipid biomarker-based approach to evaluate microbial community changes across the Permo/Triassic (P/Tr) boundary at Meishan in South China. We present here a biomarker stratigraphic record showing episodic microbial changes coupled with a high-resolution record of invertebrate mass extinction. Variation in the microbial community structure is characterized by the 2-methylhopane (2-MHP) index (a ratio of the abundance of cyanobacterial biomarkers to more general bacterial biomarkers). Two episodes of faunal mass extinction were each preceded by minima in the 2-MHP index, followed by strong maxima, likely reflecting microbial responses to the catastrophic events that caused the extinction and initiated ecosystem changes. Hence, both cyanobacterial biomarker and invertebrate fossil records provide evidence for two episodes of biotic crisis across the P/Tr boundary.

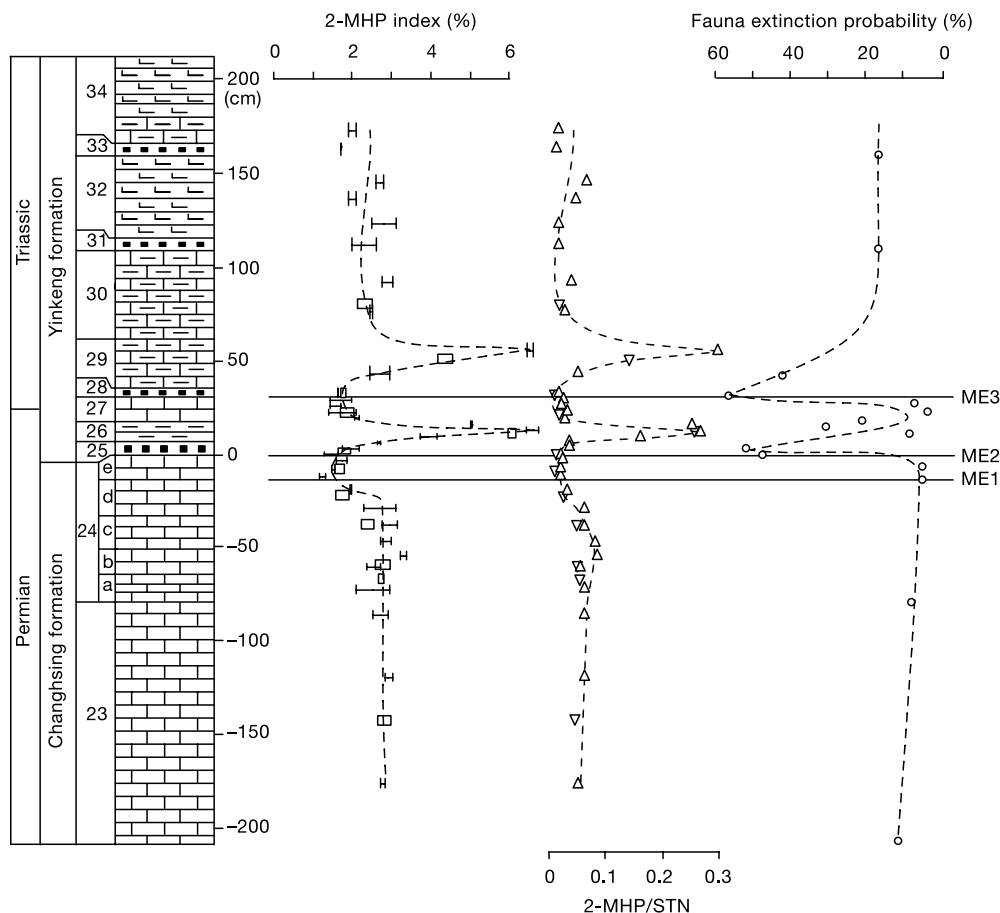
In bacteria, bacteriohopanepolyols (BHPs) are amphiphilic biochemicals that serve regulatory and rigidifying functions in cell membranes. It has been demonstrated that the cyanobacterial lineage currently comprises the only known quantitatively significant source of 2-methyl-BHPs<sup>4–6</sup>, such that the presence of abundant 2 $\alpha$ -methylhopanes (2-MHPs) derived from 2-methyl-BHP precursors is believed to be characteristic of cyanobacteria<sup>4</sup>. This is particularly true for 2-methyl-BHPs containing more than 32 carbon atoms, which have not been detected in organisms other

than cyanobacteria<sup>4</sup>. Thus, the 2-MHP index, which is the ratio of 2-MHPs to hopanes lacking additional methyl substituents (HPs, derived from a range of bacteria, including cyanobacteria<sup>7</sup>), provides insight into the cyanobacterial component of the bacterial population.

We have conducted a survey of microbial biomarkers in sedimentary rocks spanning the P/Tr boundary (which is characterized by the most severe biotic crisis of the Phanerozoic) and compared these biomarkers to faunal extinction patterns. We targeted the Meishan sections in South China for several reasons. First, they contain well established, high-resolution faunal stratigraphic distributions that are well dated<sup>8</sup>, thereby allowing comparison of the fossil biomarker record with macro- and micro-fossil records. Second, they are the most extensively studied P/Tr boundary sections anywhere in the world<sup>8</sup> and are where the Global Stratotype Section and Point (GSSP) was recently defined<sup>9</sup>. Third, they are the location at which the rapid nature of biotic crises has been proposed<sup>8</sup>. Finally, they possess organic matter that is relatively thermally immature, as indicated by  $T_{\text{max}}$  values of 424–466 °C (temperature of maximum pyrolysis yield in Rock-Eval pyrolysis; see Supplementary Table 1), making it possible to investigate biomarkers.

The 2-MHP indices were determined by gas chromatography/mass spectrometry (GC/MS) of the saturated hydrocarbon fraction of solvent extracts of the sedimentary rocks. 2-MHP indices range from 1.2% to 6.6%, consistent with previously reported values for the Paleozoic<sup>4</sup>. In general, values are between 2.0% and 3.3%, but the record is characterized by two minima (with values of 1.2–1.7%) at bed 24e (extending to the bottom of bed 25) and beds 27–28; each of these is followed by a dramatic maximum (6.6%) at beds 26 and 29, respectively (Fig. 1). These maxima indicate at least two major shifts in the cyanobacterial community, either through changes in the cyanobacterial species composition (because not all cyanobacteria biosynthesize 2-methyl-BHPs) or an increase in the size of the total cyanobacterial community. Although an alternative explanation is that changes in the 2-MHP depth profile record varying abundances of a less common 2-MHP producer, this is unlikely because these episodes are also associated with enhanced relative abundances of more diagnostic 2-MHPs with extended side chains, including the C<sub>32</sub> 22R and 22S isomers<sup>4,10</sup>.

Other than changes in cyanobacterial community structure, the increase in 2-MHP indices could be caused by differences in either thermal maturity or lithology/facies. The former is unlikely given the narrow stratigraphic range, confirmed by the weak correlation



**Figure 1** Profiles of the lithology, 2-MHP indices, the ratio of 2-MHP to steranes (2-MHP/STN) and the extinction probability of marine invertebrates across the P/Tr boundary at the Meishan sections in Zhejiang, South China. The lithology of the studied profiles near the P/Tr boundary includes lime mudstones (beds 23, 24), pale-coloured, alternated ash clay beds (beds 25, 28, 31, 33), a laminated organic-rich calcareous claystone (bed 26), a lime mud bed (bed 27), and grey organic-rich shales, pale marls or muddy limestone (beds 29, 30, 32, 34). The high-resolution 2-MHP indices are from Meishan Section B and C (the error bars reflect the error:  $2\sigma$  of  $>3$  measurements). The ratio of 2-MHP to steranes is the abundance of C<sub>31</sub> 2-MHP relative to C<sub>27</sub>–C<sub>29</sub> regular steranes (with  $1\sigma$  less than the width of the triangular points) from section B (upright

triangles) and C (inverted triangles). The extinction probability of individual horizons was calculated (by dividing numbers of extinction species by the total number of species excluding originating species) from previously reported high-resolution species-occurrence records<sup>8</sup>. Mass extinctions ME1, 2 and 3 show the start horizons of the three previously proposed extinction events<sup>12,27</sup>. ME1 is associated with the loss of the latest Permian reef ecosystem in South China, which corresponds to bed 24e at Meishan. ME2 is the disappearance of most benthos at beds 25–26, and ME3 is the disappearance of the final Permian brachiopods at beds 28–29. Numerical data can be found in Supplementary Tables 1 and 2.



( $r^2 = 0.04$ ) between  $T_{\max}$  values and 2-MHP indices through the section (Fig. 2). Moreover, facies changes cannot explain the observed variation; although proposed water-column deepening at beds 24e, 25 and 28 (refs. 11–13) do coincide with 2-MHP minima, the maxima in 2-MHP indices are not associated with lithological or facies variation. In fact, 2-MHP indices show no significant correlation with total organic carbon (TOC) content ( $r^2 = 0.06$ ), and vary within individual beds (for example, beds 26 and 29; Fig. 2). Thus, the observed changes in 2-MHP indices probably reflect changes in the bacterial community structure. Because the 2-MHP index is a ratio, changes in values could reflect either changes in cyanobacterial (2-MHP) or total bacterial (HP) abundances. However, TOC-normalized abundances of 2-MHP and the ratio of 2-MHP to steranes (biomarkers of eukaryotes, primarily algae) have maxima coincident with those of the 2-MHP indices (Fig. 1), indicating that changes in the cyanobacterial populations underlie the observed trends.

In order to understand the relationship between microbial change and faunal variation, we counted the numbers of faunal species becoming extinct, together with the survivors and originating species at each horizon across the P/Tr boundary of the Meishan sections (Fig. 1 and Supplementary Table 2). This allowed calculation of extinction probability (by dividing the numbers of extinction species by the total numbers of species, excluding the originating species, at a given level) for each stratigraphic horizon<sup>8</sup>.

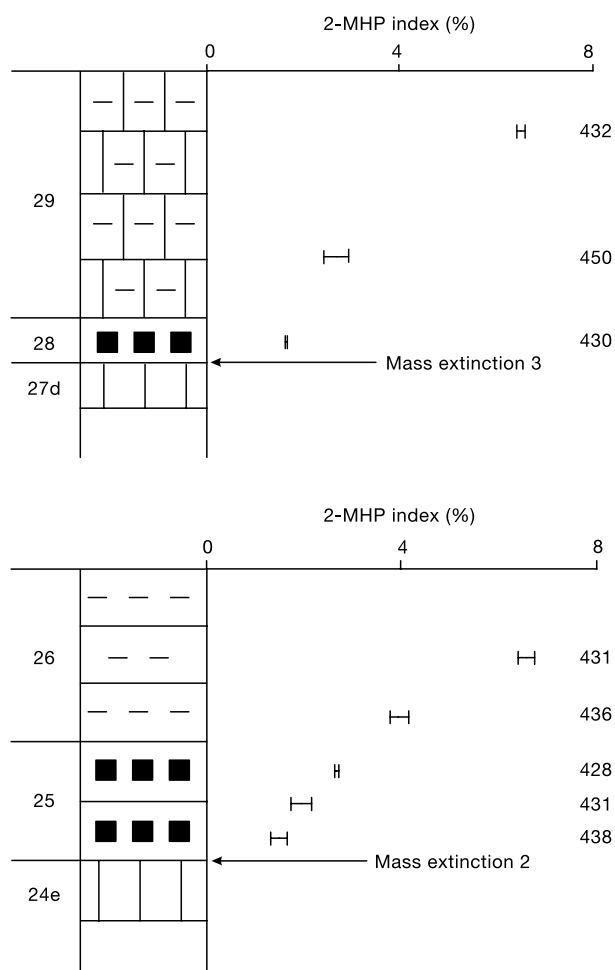
The record revealed two pulses of faunal extinction at beds 25 and 28–29 (Fig. 1 and Supplementary Table 2). It is likely that these correspond, respectively, to the previously proposed mass extinctions 2 and 3 documented in several sections in southern China<sup>12</sup>. This correlation between the reported multiple mass extinctions and specific horizons at the Meishan sections has previously been inferred<sup>8</sup>. Critically, the extinction probability and the 2-MHP indices are stratigraphically coupled, with minima in 2-MHP indices coinciding with, or slightly preceding, the times of maximum extinction, while maxima in 2-MHP indices clearly lag extinction.

Based on the above, it appears that faunal and microbial turnover were closely related. This probably reflects: (i) distinct but coincidental responses to external environmental forcings, for example, nutrient fluctuations; (ii) changes in both microbial and faunal populations, reflecting their ecological coupling (for example, grazing pressure or space competition); or (iii) a combination of the two. Moreover, the change in planktonic cyanobacterial populations at Meishan appears to coincide with the expansion of cyanobacterial benthos in a vast area of the eastern Tethys and Panthalassa inferred from petrological investigation of microbialites<sup>1</sup>. These authors proposed that benthic invertebrates limited benthic cyanobacterial populations by co-opting space and grazing, such that extinction of benthic invertebrates allowed the cyanobacteria to expand<sup>1,2</sup>. However, the former mechanism probably does not apply to the Meishan sections because photosynthetic microbial mats are not expected in the shelf margin to deeper-water slope facies (above bed 24e; refs 8, 14). Given this, the Meishan 2-MHP indices probably record changes in the planktonic community, reflecting changes in nutrient status<sup>15</sup> or grazing pressures<sup>16</sup>.

The size of phytoplankton populations, including cyanobacterial members, can be suppressed by grazing<sup>15–17</sup>. Significantly, grazing pressure in the wake of the P/Tr extinction could have been much reduced in the photic zone, causing the expansion of cyanobacterial populations. The end-Permian faunal extinctions have apparently selected for planktotrophs<sup>16,18</sup>, which are dominant in warm, shallow, low-latitude waters and feed on phytoplankton<sup>18</sup>; this has previously been invoked as a mechanism for photoautotroph population expansions at the P/Tr boundary<sup>15,19</sup>. The association of enhanced 2-MHP abundances relative to steranes with the two episodes of faunal mass extinction (Fig. 1 and Supplementary Table 1) suggests that reduced grazing might specifically favour cyanobacterial expansion relative to planktonic algae.

The temporal lags in geological magnitude between faunal extinction and maximum cyanobacterial expansion are longer than the delays in population changes between predator and prey in modern ecosystems<sup>20</sup>. However, previous workers have shown that trophic relationships, including those between phyto- and zooplankton, can develop over extended time periods in the wake of an extinction event<sup>21</sup>. For example, a lag between microzooplankton and microphytoplankton diversity changes, linked through trophic relationships, was recently identified in Middle Ordovician strata<sup>22</sup>; likewise, a proliferation of acritarchs occurs 1–2 m above the P/Tr faunal mass extinction in strata from Tesero, Italy<sup>23</sup>. Our P/Tr record is also similar to the photoautotroph response to the Cretaceous/Tertiary (K/T) extinction event, which apparently resulted in temporarily low marine productivity followed by enhanced productivity a few hundred thousand years later<sup>21</sup>.

Although a decrease in grazing pressure is a likely explanation for the cyanobacterial expansion observed here, a contribution from changes in nutrient dynamics cannot be excluded. Cyanobacterial proliferation has been observed in response to enhanced nutrient inputs in modern aquatic environments<sup>24</sup>. Elevated nutrient levels in the end-Permian seas and oceans could have been caused by intense weathering throughout Pangea, as documented by end-



**Figure 2** Plots showing the increase in 2-MHP indices (error bars,  $2\sigma$ ) within individual beds after the faunal mass extinctions 2 and 3, together with the  $T_{\max}$  values (numbers on right-hand side) indicating the maturity of the organic matter. Numerical data can be found in Supplementary Tables 1 and 2.

Permian pedoliths<sup>25</sup> and changed geomorphology as a result of rooted vegetation loss<sup>26</sup>.

Regardless of the detailed mechanisms, the striking aspect of the Meishan 2-MHP profile is the presence of two distinct maxima within the investigated intervals. The pattern and cause of the marine mass extinction near the P/Tr boundary are still debated<sup>8,12,27,28</sup>, in part because only faunal (and a small amount of calcareous algae) records of biotic change have been developed. A stepwise, multi-phase biotic extinction proposed by others on the basis of several southern China sections<sup>12,27</sup>, has recently been challenged by a proposal of a main, sudden biotic crisis at beds 25–26 (ref. 28). However, the change in the 2-MHP index we report is equally dramatic (from minima to maxima) at both beds 25–26 and 28–29, such that a two-phase biotic crisis is supported not only by faunal extinctions, the interpretation of which has been the subject of debate, but also by substantial changes in the microbial community. The large changes in cyanobacterial population abundances revealed by the 2-MHP record, coupled with faunal mass extinction, probably indicate a stepwise and major environmental perturbation. Our study provides a window into the ecological restructuring, possibly at the base of the food chain, that occurred during this major faunal mass extinction. □

## Methods

The solvents (HPLC grade) used were distilled twice. Tools used to prepare rock samples were rinsed with methanol and dichloromethane. Glassware was cleaned with detergent (Micro), rinsed with distilled water and annealed at 450 °C for 12 h. Aluminium foils were combusted at 450 °C for 12 h. Samples were processed in batches of six, including one procedural blank to monitor laboratory contamination.

The samples for lipid analysis were taken from section B and C at Meishan<sup>9</sup>, 0.5–1.5 m back from the exposed cliff. The air-dried samples were ground to <80 mesh and 100–300 g Soxhlet-extracted with chloroform for 72 h with native copper added to remove sulphur. The extract was concentrated on a rotary evaporator under reduced pressure and transferred to a small vial. After evaporation of the remaining solvent, the total extractable lipid was weighed. The asphaltenes in the extracts were eliminated by precipitation in petroleum ether. The saturated hydrocarbons, aromatics and non-hydrocarbons were fractionated by column chromatography (silica gel 60), using sequential elution with petroleum ether, benzene and ethanol.

The saturated hydrocarbons were then analysed by GC/MS using a Hewlett-Packard 5973A MS, interfaced directly with a 6890 GC equipped with a HP-5MS fused silica capillary column (30 m × 0.25 mm inner diameter; 0.25 µm film thickness). The operating conditions were as follows: temperature ramped from 70 to 280 °C at 3 °C min<sup>-1</sup> and held at 280 °C for 20 min with He as the carrier gas; the ionization energy of the mass spectrometer was set at 70 eV; the scan range was *m/z* 50–550. 2-MHP indices were calculated from the peak area of the compounds in mass chromatograms of *m/z* 191 (C<sub>30</sub>) for αβ-hopane and *m/z* 205 (C<sub>31</sub>) for its 2α-methyl analogue (similar profiles are generated if 2-MHP indices are calculated using higher molecular mass 2-MHPs and HPs). These are expressed as the 2-MHP index = (205/191) × 100. Sterane abundances used to calculate 2-MHP/sterane ratios were measured from the peak areas of the *m/z* 217 mass chromatograms. 2-MHPs were not detected in the procedural blanks using GC/MS analysis.

Received 26 August 2004; accepted 18 January 2005; doi:10.1038/nature03396.

- Lehrmann, D. J. *et al.* Permian–Triassic boundary sections from shallow-marine carbonate platforms of the Nanpanjiang Basin, south China: Implications for oceanic conditions associated with the end-Permian extinction and its aftermath. *Paleosol* **18**, 138–152 (2003).
- Sheehan, P. M. & Harris, M. T. Microbialite resurgence after the Late Ordovician extinction. *Nature* **430**, 75–78 (2004).
- Knoll, A. H. & Walter, M. W. Latest Proterozoic stratigraphy and Earth history. *Nature* **356**, 673–678 (1992).
- Summons, R. E., Jahnke, L. L., Hope, J. M. & Logan, G. A. 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature* **400**, 554–557 (1999).
- Simonin, P., Jurgens, U. J. & Rohmer, M. Bacterial triterpenoids of the hopane series from the prochlorophyte *Prochlorothrix hollandica* and their intracellular localization. *Eur. J. Biochem.* **241**, 865–871 (1996).
- Rohmer, M., Bouvier-Nave, P. & Ourisson, G. Distribution of hopanoid triterpenes in prokaryotes. *J. Gen. Microbiol.* **130**, 1137–1150 (1984).
- Rohmer, M., Bissert, P. & Neunlist, S. in *Biological Markers in Sediments and Petroleum* (eds Moldovan, J. M., Albrecht, P. & Philp, R. P.) 1–17 (Prentice Hall, New Jersey, 1992).
- Jin, Y. G. *et al.* Pattern of marine mass extinction near the Permian–Triassic boundary in South China. *Science* **289**, 432–436 (2000).
- Yin, H., Zhang, K., Tong, J., Yang, Z. & Wu, S. The Global Stratotype Section and Point (GSSP) of the Permian–Triassic boundary. *Episodes* **24**, 102–114 (2001).
- Brocks, J. J., Buick, R., Summons, R. E. & Logan, G. A. A reconstruction of Archean biological diversity based on molecular fossils from the 2.78 to 2.45 billion-year-old Mount Bruce Supergroup, Hamersley Basin, Western Australia. *Geochim. Cosmochim. Acta* **67**, 4321–4335 (2003).
- Wignall, P. B. & Hallam, A. Griesbachian (Earliest Triassic) palaeoenvironmental changes in the Salt

Range, Pakistan and Southeast China and their bearing on the Permian–Triassic mass extinction. *Paleogeogr. Palaeoclimatol. Palaeoecol.* **102**, 215–237 (1993).

- Yin, H. *et al.* The Palaeozoic–Mesozoic Boundary, Candidates of Global Stratotype Section and Point of the Permian–Triassic Boundary (China Univ. of Geosciences Press, Wuhan, 1996).
- Lai, X., Wignall, P. & Zhang, K. Palaeoecology of the conodonts *Hindeodus* and *Clarkina* during the Permian–Triassic transitional period. *Paleogeogr. Palaeoclimatol. Palaeoecol.* **171**, 63–72 (2001).
- Yin, H. & Tong, J. Multidisciplinary high-resolution correlation of the Permian–Triassic boundary. *Paleogeogr. Palaeoclimatol. Palaeoecol.* **143**, 199–212 (1998).
- Hallam, A. & Wignall, P. *Mass Extinctions and their Aftermath* (Oxford Univ. Press, Oxford, 1997).
- Valentine, J. W. & Jablonski, D. Mass extinctions: sensitivity of marine larval types. *Proc. Natl Acad. Sci. USA* **83**, 6912–6914 (1986).
- Dionisio-Pires, L. M., Jonker, R. R., Van Donk, E. & Laanbroek, H. J. Selective grazing by adults and larvae of the zebra mussel (*Dreissena polymorpha*): application of flow cytometry to natural seston. *Freshwat. Biol.* **49**, 116–126 (2004).
- Erwin, D. H. The end-Permian mass extinction. *Annu. Rev. Ecol. Syst.* **21**, 495–516 (1990).
- Racki, G. Silica-secreting biota and mass extinctions: survival patterns and processes. *Paleogeogr. Palaeoclimatol. Palaeoecol.* **154**, 107–132 (1999).
- Keeling, M. J., Wilson, H. B. & Pacala, S. W. Reinterpreting space, time lags, and functional responses in ecological models. *Science* **290**, 1758–1761 (2000).
- Erwin, D. H. Lessons from the past: biotic recoveries from mass extinction. *Proc. Natl Acad. Sci. USA* **98**, 5399–5403 (2001).
- Vecoli, M. & Le Herisse, A. Biostratigraphy, taxonomic diversity and patterns of morphological evolution of Ordovician acritarchs (organic-walled microphytoplankton) from the northern Gondwana margin in relation to palaeoclimatic and palaeogeographic changes. *Earth-Sci. Rev.* **67**, 267–311 (2004).
- Rampino, M. R. & Adler, A. C. Evidence for abrupt latest Permian mass extinction of foraminifera: results of tests for the Signor-Lipps effect. *Geology* **26**, 415–418 (1998).
- Mastin, B. J., Rodger, J. H. Jr & Deardorff, T. L. Risk evaluation of cyanobacteria-dominated algal blooms in a North Louisiana reservoir. *J. Aquatic Ecosyst. Stress Recovery* **9**, 103–114 (2002).
- Retallack, G. J. Search for evidence of impact at the Permian–Triassic boundary in Antarctica and Australia. *Geology* **26**, 979–982 (1998).
- Ward, P. D., Montgomery, D. R. & Smith, R. Altered river morphology in South Africa related to the Permian–Triassic extinction. *Science* **289**, 1740–1743 (2000).
- Yang, Z. Y. *et al.* *Permian–Triassic Events of South China* (Geological Publishing House, Beijing, 1993).
- Erwin, D. H., Bowring, S. A. & Jin, Y. in *Special Paper 356, Catastrophic Events and Mass Extinctions: Impacts and Beyond* (eds Koeberl, C. & MacLeod, K. G.) 363–383 (Geological Society of America, Boulder, 2002).

Supplementary Information accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank F. Yang, S. Wu, G. Zhu, J. Huang, Y. Yi and J. Yu for field assistance, L. Lu, D. Jiao and X. Huang for sample processing, and M. Benton for constructive comments on the manuscript. The authors thank I. D. Bull and R. Berstan for their technical assistance and the NERC for funding the Bristol node of the Life Sciences Mass Spectrometry Facility (<http://www.chm.bris.ac.uk/lsmf/index.html>). This work was supported by National Natural Science Foundation (H. Y.), NCET programme of Ministry of Education of China (S. X.) and the Lab of Bio- and Environmental Geology in China University of Geosciences.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to R.D.P. ([r.d.pancost@bristol.ac.uk](mailto:r.d.pancost@bristol.ac.uk)).

## Affinities of ‘hyopsodontids’ to elephant shrews and a Holarctic origin of Afrotheria

Shawn P. Zack<sup>1</sup>, Tonya A. Penkrot<sup>1</sup>, Jonathan I. Bloch<sup>2</sup> & Kenneth D. Rose<sup>1</sup>

<sup>1</sup>Center for Functional Anatomy and Evolution, The Johns Hopkins University School of Medicine, 1830 East Monument Street, Baltimore, Maryland 21205, USA

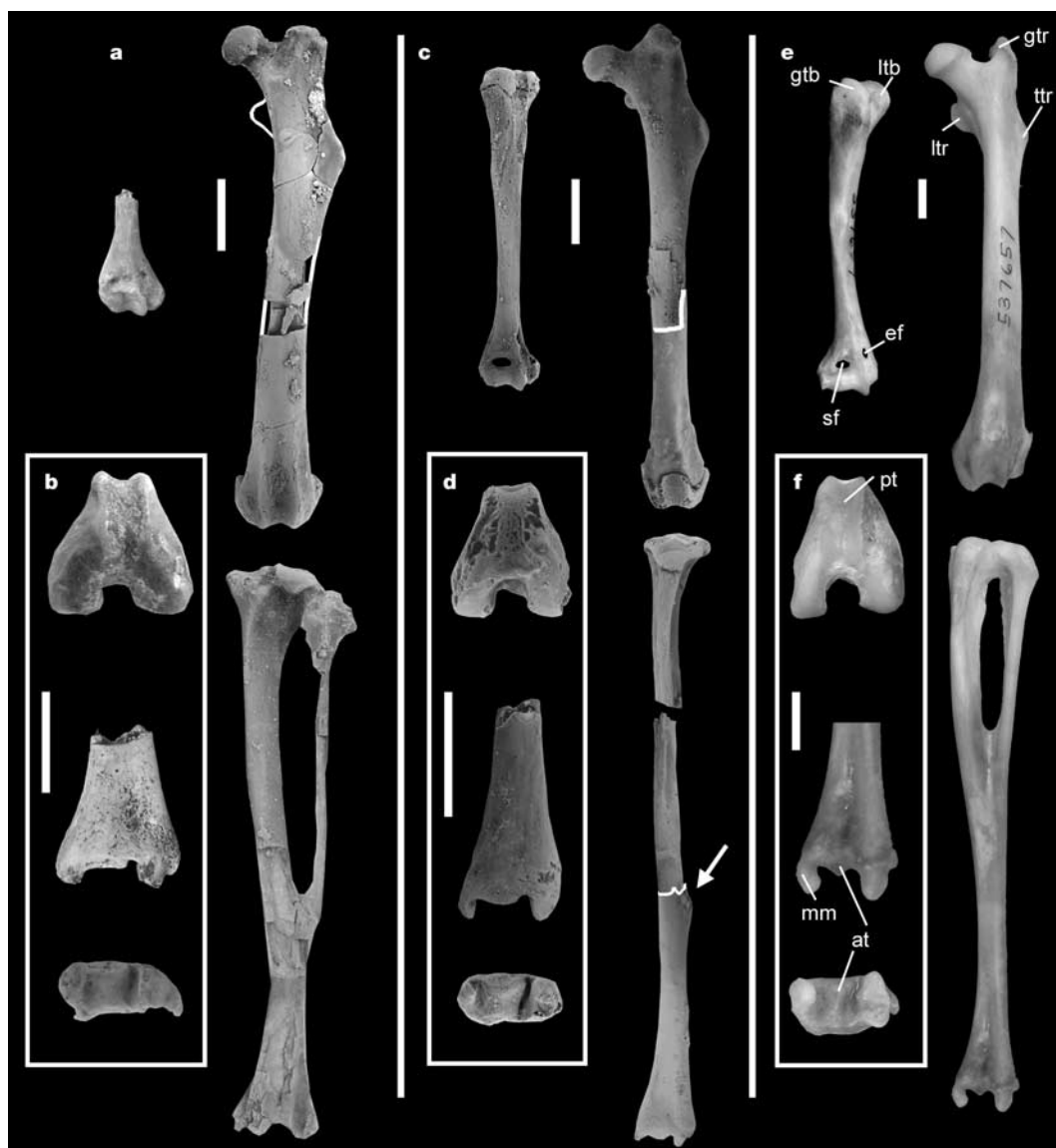
<sup>2</sup>Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611–7800, USA

Macroscelideans (elephant shrews or sengis) are small-bodied (25–540 g), cursorial (running) and saltatorial (jumping), insectivorous and omnivorous<sup>1</sup> placental mammals represented by at least 15 extant African species classified in four genera<sup>2</sup>. Macroscelidea is one of several morphologically diverse but predominantly African placental orders classified in the superorder

Afrotheria by molecular phylogeneticists<sup>3,4</sup>. The distribution of modern afrotheres, in combination with a basal position for Afrotheria within Placentalia and molecular divergence-time estimates, has been used to link placental diversification with the mid-Cretaceous separation of South America and Africa<sup>4</sup>. Morphological phylogenetic analyses do not support Afrotheria<sup>5–7</sup> and the fossil record favours a northern origin of Placentalia<sup>8</sup>. Here we describe fossil postcrania that provide evidence for a close relationship between North American Palaeocene–Eocene apheliscine ‘hyopsodontid’ ‘condylarths’ (early ungulates or hoofed mammals) and extant Macroscelidea. Apheliscine postcranial morphology is consistent with a relationship to other ungulate-like afrotheres (Hyracoidea,

Proboscidea) but does not provide support for a monophyletic Afrotheria. As the oldest record of an afrothere clade, identification of macroscelidean relatives in the North American Palaeocene argues against an African origin for Afrotheria, weakening support for linking placental diversification to the break-up of Gondwana.

Molecular analyses that place Macroscelidea within Afrotheria also include tenrecid and chrysochlorid insectivores, paenungulates (hyracooids, proboscideans and sirenians), and tubulidentates<sup>3,4</sup> in this clade, a diverse assemblage of predominantly African placentals. Morphologists have proposed a relationship for Macroscelidea to either Glires (rodents and rabbits)<sup>9–11</sup> or to Ungulata<sup>12–15</sup>, the latter also supported by behavioural observations<sup>2</sup>. Morphologists



**Figure 1** Comparison of apheliscine and macroscelidean long bones. Long bones of *Apheliscus* (**a, b**), *Haplomylus* (**c, d**), and the macroscelidean *Rhynchocyon* (**e, f**). Elements in **a, c** and **e** (clockwise from top left): right humerus, left femur and left fused tibia and fibula in anterior view except *Apheliscus* tibia-fibula (in posterior view). Elements in **b, d** and **f** (top to bottom): left distal femur in distal view, distal tibia-fibula in anterior view, and distal tibia-fibula in distal view. Some elements are reversed for clarity. Scale bars, 5 mm. Arrow in **c** indicates the most proximal point of fusion of the tibia and fibula. at, anterior tubercle; ef, entepicondylar foramen; gtb, greater tuberosity; gtr, greater trochanter; ltb, lesser tuberosity; ltr, lesser trochanter; mm, medial malleolus; pt, patellar

groove; sf, supratrochlear foramen; ttr, third trochanter. Specimens illustrated (United States National Museum (USNM); Department of Paleobiology (**a–d**) and Department of Mammalogy (**e–f**)): **a**, 493903 (humerus), 488326 (femur, ltr reconstructed from 525593), 495051 (tibia-fibula); **b**, 493819 (femur), 493903 (tibia-fibula); **c**, 513512 (humerus), 513057 (proximal femur), 513140 (distal femoral shaft), 513173 (distal femoral epiphysis), 513245 (proximal tibia), 513868 (tibia-fibula); **d**, 513173 (femur), 513239 (tibia-fibula); **e–f**, 537657. See Supplementary Information for a full list of specimens examined.



include artiodactyls, perissodactyls, paenungulates, possibly tubulidentates, and extinct groups, particularly the paraphyletic basal order 'Condylarthra'<sup>9,16,17</sup>, in Ungulata. The fossil record of macroscelideans extends to the African late early Eocene<sup>12</sup>, and these dental fossils suggest affinities to the polyphyletic<sup>15,18</sup> 'condylarth' family 'Hyopsodontidae'. However, *Hyopsodus*, the only skeletally well-known 'hyopsodontid', differs markedly from macroscelideans<sup>19</sup>.

A new dentally associated skeleton of *Apheliscus chydaeus* (USNM 525597) includes portions of all long bones and several foot bones, allowing identification of other isolated bones from Willwood Formation quarries (early Eocene, Bighorn Basin, Wyoming). Direct associations of dental and postcranial material remain unknown for *Haplomylus*, but a convincing case for the reassociation of isolated postcranial elements can be made on the basis of size, relative abundance, stratigraphic distribution and especially morphologic similarity to those of *Apheliscus*. Dental morphology has been used to link *Apheliscus* and *Haplomylus* (referred to informally as apheliscines), along with other 'Hyopsodontidae'<sup>20,21</sup>.

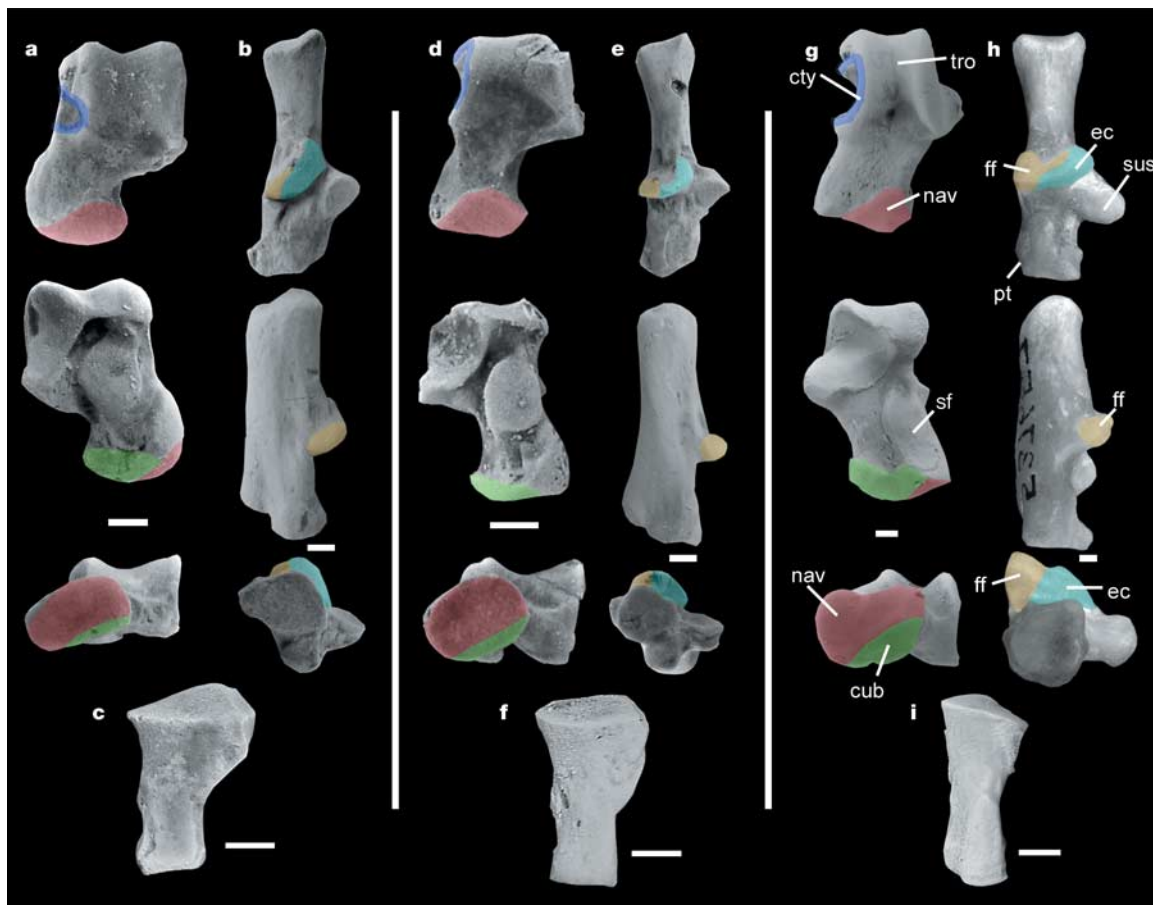
Apheliscine postcrania indicate a similar body mass to the extant forest elephant shrew *Petrodromus tetradactylus* (160–280 g<sup>22</sup>) and specializations for rapid terrestrial locomotion, as in extant cursorial and saltatorial mammals (see Supplementary Information for modern taxa examined). The combination of

specializations in apheliscines is uniquely similar to living and extinct macroscelideans, suggesting a close phylogenetic relationship.

In the forelimb, the distal humerus has a weak supinator crest and a reduced medial epicondyle. The olecranon fossa is deep (*Apheliscus*) or perforate (*Haplomylus*), indicating capacity for full extension at the elbow. The capitulum is ovoid and the trochlea has a sharply distally projecting medial rim, whereas the ovoid proximal radius has a flat ulnar articulation, indicating little or no ability to supinate the forearm (Fig. 1a, c).

Features of the femur indicate specialization for rapid terrestrial locomotion. Proximally, the greater trochanter is even with (*Apheliscus*) or slightly higher (*Haplomylus*) than the small femoral head. The femoral neck is short and distinct. The lesser trochanter is directed posteromedially, whereas the third trochanter is strong and proximally positioned. Distally, the femur is anteroposteriorly deep, with anteriorly extensive condylar articular surfaces, permitting full knee extension, whereas the patellar groove is narrow, deep, elevated and proximally extensive.

Specializations for cursoriality are also found in the crus. The proximal tibia has a sharp cnemial crest that extends only one-quarter (*Haplomylus*) to one-third (*Apheliscus*) the length of the bone, indicating capability for rapid extension at the knee. The gracile shafts of the tibia and fibula are completely synostosed over the distal one-third (*Apheliscus*) to one-half (*Haplomylus*) of their



**Figure 2** Comparison of apheliscine and macroscelidean tarsals. Left astragali (**a, d, g**), right calcanei (**b, e, h**) and cuboids (**c, f, i**) of *Apheliscus* (**a–c**), *Haplomylus* (**d–f**), *Rhynchocyon* (**g, h**) and *Petrodromus* (**i**). Astragali illustrated in dorsal (top), ventral (middle) and distal (bottom) views. Calcanei illustrated in dorsal (top), lateral (middle) and distal (bottom) views. Cuboids illustrated in dorsal view. Some elements are reversed for clarity. Scale bars, 1 mm. cty, cotylar fossa (dark blue); cub, cuboid facet (green); ec, ectal

facet (light blue); ff, fibular facet (yellow); nav, navicular facet (red); pt, peroneal tubercle; sf, sustentacular facet (astragalar); sus, sustentacular facet (calcaneal); tro, trochlea. Specimens illustrated (USNM; Department of Paleobiology (**a–f**) and Department of Mammalogy (**g–i**)): **a**, 521791; **b**, 521789; **c**, 513632; **d**, 493901; **e–f**, 537657.

length. The tibia is longer than the femur, an estimation based on unassociated elements (Fig. 1).

In the tarsus, the astragalar trochlea is deeply grooved and lacks an astragalar foramen. The astragalar neck is elongate. The axis of curvature of the navicular facet on the astragalar head indicates predominance of parasagittal movements at the transverse tarsal joint. Characteristic of digitigrade mammals (in which weight is supported on distal phalangeal segments) such as modern dogs, cats and rabbits, the calcaneal tuber is elongate and gracile. The body of the calcaneum is elongate, and the peroneal tubercle is small (*Apheliscus*) or almost absent (*Haplomylus*), and is situated at the distal margin of the calcaneum (Fig. 2). On the cuboid, the metatarsal facet is flat and faces distally (not distolaterally), suggesting a columnar metatarsus (again implying digitigrady).

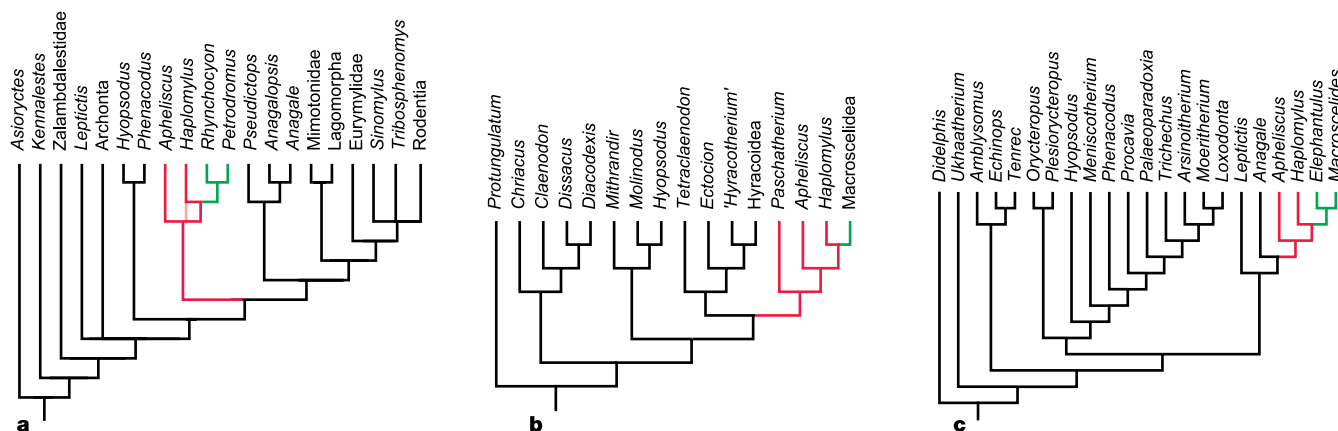
The characteristics described above are present in diverse extant cursorial and saltatorial mammals, but the combination in apheliscines occurs only in macroselideans. Additional postcranial features infrequently encountered among mammals further support a special relationship between apheliscines and macroselideans. In *Apheliscus*, the proximal fibula is enlarged and contributes to the posterolateral corner of the knee joint, a trait elaborated on in macroselideans. The proximal fibula is unknown in *Haplomylus*, but facets on the tibia indicate further expansion onto the lateral margin of the lateral tibial condyle. In apheliscines, the medial malleolus of the distal tibia is enlarged to form a laterally recurved hook-like process, whereas the central portion of the deeply grooved distal tibial articular surface slopes anteroventrally to form a distinctive anterior tubercle. Both of these features are shared with macroselideans, as is the presence of a well-developed cotylar fossa—larger in *Haplomylus* than in *Apheliscus*—that articulates with the tibial malleolus on the medial aspect of the astragalar body. A large cuboid facet is present on the astragali of both apheliscines, and the boundary between the cuboid and navicular facets is sharp in *Haplomylus*, another similarity to macroselideans. In *Haplomylus*, as in macroselideans, the angle between the proximal and distal portions of the calcaneal ectal facet is sharp, restricting mobility between the astragalus and calcaneum. Apheliscines resemble macroselideans in having a large calcaneal fibular facet whose axis of curvature parallels that of the astragalar trochlea (Fig. 2). Finally, apheliscines share an elongate cuboid with macroselideans (Fig. 2c,f,i). The combination of these distinctive features is found only in apheliscines and macroselideans, providing

particularly powerful evidence for a close relationship, although no individual feature is unique. This is strengthened by previous phylogenetic hypotheses, based on dental anatomy, that some 'hyposodontids'—particularly *Haplomylus*—and macroselideans are closely related<sup>13,14</sup>.

In view of proposed relationships of Macroselidea to Glires (in the superorder Anagalida), Ungulata (as traditionally defined to include paenungulates), or Afrotheria, we tested the hypothesized macroselidean affinities of *Apheliscus* and *Haplomylus* using three character-taxon matrices: two previously published that extensively sample Anagalida and Afrotheria, respectively<sup>7,11</sup>, and a new matrix that samples basal members of Ungulata (Supplementary Information). Phylogenetic analyses of all three matrices, using the parsimony ratchet algorithm of NONAv2.0 implemented using Winclada(BETA)v0.9.9, place apheliscines on the macroselidean stem, with *Haplomylus* as the sister taxon to extant African macroselideans (Fig. 3). Results from our new analysis indicate that the European lousinine 'hyposodontid' *Paschatherium* (known from dentitions and tarsal bones) is the sister taxon to the clade that includes apheliscines and extant African Macroselidea. *Hyposodus* is not closely related to extant Macroselidea or apheliscines in any of the three analyses, supporting a polyphyletic 'Hyposodontidae'.

Identification of a relationship between apheliscines and macroselideans has direct significance to a discussion of the higher-level affinities of Macroselidea. A close relationship between Glires and macroselideans is not supported by dental (apheliscines and early macroselideans are more ungulate-like<sup>12–15</sup>) or postcranial morphology (apheliscines lack anagalidan characteristics such as a posterior process on the distal tibia). Because most character evidence recently cited in support of Anagalida is cranial<sup>9,10</sup>, we cannot definitively reject anagalidan affinities for Macroselidea, but the known morphology of *Apheliscus* and *Haplomylus* argues against such a relationship.

On the basis of dental morphology<sup>16–18,21</sup> most authors have placed *Haplomylus* and *Apheliscus* in the basal ungulate order 'Condylarthra', and it is not surprising that apheliscine dentitions are more congruent with ungulate affinities for Macroselidea. The general postcranial resemblance of apheliscines to many ungulates, particularly artiodactyls, perissodactyls and phenacodontid 'condylarths', probably reflects similar adaptations for rapid terrestrial locomotion<sup>23</sup>. The lack of more distinctive derived features in common with any of those groups (for example, the double-pulley



**Figure 3** Phylogenetic relationships of apheliscines. **a**, Strict consensus of 16 most parsimonious trees from reanalysis of the ref. 11 character-taxon matrix with some characters ordered (length, L: 979; consistency index, CI: 39; retention index, RI: 77). For visual simplicity, some clades are collapsed into single terminal taxa. Faint red line: position of *Haplomylus* in strict consensus of twelve most parsimonious trees with all characters unordered (L: 926; CI: 41; RI: 76). **b**, Single most parsimonious tree from

analysis of basal ungulates with all characters unordered (L: 180; CI: 38; RI: 47) or with some characters ordered (L: 181; CI: 38; RI: 48). **c**, Single most parsimonious tree derived from reanalysis of the ref. 7 character-taxon matrix with all characters unordered and only afrotheres and their potential extinct relatives included (L: 380; CI: 45; RI: 58). In all trees, macroselideans are indicated in green, whereas red indicates apheliscines (and *Paschatherium* in **b**).

astragalus of artiodactyls) indicates that these similarities represent homoplasies rather than synapomorphies.

The cotylar fossa (and associated enlarged medial malleolus), present in apheliscines and macroselideans, is shared with hyracoids, proboscideans, tubulidentates, and their possible extinct relatives *Meniscotherium* and *Plesiorcyteropus*<sup>24–26</sup>. Living members of this assemblage of mammals have been classified in Ungulata by morphological systematists, but in Afrotheria by molecular systematists. Presence of a cotylar fossa does not seem to be associated with any particular mode of locomotion, because it occurs in fossorial (Tubulidentata, *Plesiorcyteropus*) and scansorial (Hyracoidea) mammals, as well as in the cursorial/saltatorial macroselideans. The cotylar fossa is also present in cercopithecoid primates and macropodid marsupials<sup>25,27</sup>, and is therefore not unique to potential afrotheres, but it is absent in most mammals, including other cursorial and saltatorial mammals. The broad functional but restricted taxonomic distribution of the cotylar fossa makes it a potentially significant indicator of phylogenetic affinities and argues for a relationship of apheliscines and macroselideans to hyracoids, proboscideans and tubulidentates. The combination of dental and tarsal characters in apheliscines is most consistent with a relationship to Hyracoidea and Proboscidea, regardless of whether these taxa are classified in Afrotheria or a monophyletic Ungulata.

Identification of North American sister taxa to Macroselidea suggests a North American origin for the order and a non-African origin for Afrotheria. Present evidence indicates that the restriction of Macroselidea to Africa from the middle Eocene to the present is either relictual or (more likely) indicative of Palaeocene or Eocene dispersal to Africa. Therefore, the subsequent distribution of Macroselidea no longer supports an African origin. Because Holarctic Palaeocene apheliscines and *Paschatherium* potentially represent the oldest known afrotheres, a non-African origin or relictual distribution is also implied for Afrotheria, a supposition supported by early non-African records of sirenians<sup>28</sup>, proboscideans<sup>29</sup> and possibly tubulidentates<sup>30</sup>. Neither possibility supports claims that Placentalia originated in Gondwana, which are based partly on the assumption of an African origin for Afrotheria<sup>4</sup>. This conclusion agrees with that of a recent phylogenetic analysis combining morphologic and molecular data that suggested paenungulate affinities for other North American ‘condylarths’<sup>7</sup>. □

Received 22 June 2004; accepted 11 January 2005; doi:10.1038/nature03351.

- Kingdon, J. *East African Mammals. Volume IIA: Insectivores and Bats* (Univ. Chicago Press, Chicago, 1974).
- Rathbun, G. B. The social structure and ecology of elephant-shrews. *Z. Tierpsychol.* **20** (suppl.), 1–77 (1979).
- Stanhope, M. J. *et al.* Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *Proc. Natl Acad. Sci. USA* **95**, 9967–9972 (1998).
- Murphy, W. J. *et al.* Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* **294**, 2348–2351 (2001).
- Asher, R. J. A morphological basis for assessing the phylogeny of the ‘Tenrecoidea’ (Mammalia, Lipotyphla). *Cladistics* **15**, 231–252 (1999).
- Whidden, H. P. Extrinsic snout musculature in Afrotheria and Lipotyphla. *J. Mammal. Evol.* **9**, 161–184 (2002).
- Asher, R. J., Novacek, M. J. & Geisler, J. H. Relationships of endemic African mammals and their fossil relatives based on morphological and molecular evidence. *J. Mammal. Evol.* **10**, 131–194 (2003).
- Archibald, J. D. Timing and biogeography of the eutherian radiation: fossils and molecules compared. *Mol. Phylogenet. Evol.* **28**, 350–359 (2003).
- Novacek, M. J. The skull of lepidictid insectivores and the higher-level classification of eutherian mammals. *Bull. Am. Mus. Nat. Hist.* **183**, 1–111 (1986).
- Novacek, M. J. & Wyss, A. Higher-level relationships of the recent eutherian orders: morphological evidence. *Cladistics* **2**, 257–287 (1986).
- Meng, J., Hu, Y. & Li, C. The osteology of *Rhombomylus* (Mammalia, Glires): implications for phylogeny and evolution of Glires. *Bull. Am. Mus. Nat. Hist.* **275**, 1–247 (2003).
- Hartenberger, J.-L. Hypothèse paléontologique sur l’origine des Macroselidea (Mammalia). *C.R. Acad. Sci. II* **302**, 247–249 (1986).
- Simons, E. L., Holroyd, P. A. & Bown, T. M. Early Tertiary elephant-shrews from Egypt and the origin of the Macroselidea. *Proc. Natl Acad. Sci. USA* **88**, 9734–9737 (1991).
- Butler, P. M. Fossil Macroselidea. *Mamm. Rev.* **25**, 3–14 (1995).
- Tabuce, R., Coiffait, B., Coiffait, P.-E., Mahboubi, M. & Jaeger, J.-J. A new genus of Macroselidea (Mammalia) from the Eocene of Algeria: a possible origin for elephant-shrews. *J. Vertebr. Paleontol.* **21**, 535–546 (2001).

- McKenna, M. C. & Bell, S. K. *Classification of Mammals Above the Species Level* (Columbia Univ. Press, New York, 1997).
- Archibald, J. D. In *Evolution of Tertiary Mammals of North America. Volume 1: Terrestrial Carnivores, Ungulates, and Ungulatlike Mammals* (eds Janis, C. M., Scott, K. M. & Jacobs, L. L.) 292–331 (Cambridge Univ. Press, Cambridge, 1998).
- Cifelli, R. L. The origin and affinities of the South American Condylarthra and early Tertiary Litopterna (Mammalia). *Am. Mus. Novit.* **2772**, 1–48 (1983).
- Gazin, C. L. A study of the Eocene condylarthran mammal *Hyopsodus*. *Smithson. Misc. Coll.* **153**, 1–90 (1968).
- Van Valen, L. M. New Paleocene insectivores and insectivore classification. *Bull. Am. Mus. Nat. Hist.* **135**, 217–284 (1967).
- Rose, K. D. The Clarkforkian land-mammal age and mammalian faunal composition across the Paleocene-Eocene boundary. *Univ. Mich. Pap. Paleontol.* **26**, 1–197 (1981).
- Jennings, M. R. & Rathbun, G. B. *Petrodromus tetradactylus*. *Mamm. Species* **682**, 1–6 (2001).
- Rose, K. D. Postcranial skeletal remains and adaptations in early Eocene mammals from the Willowood Formation, Bighorn Basin, Wyoming. *Spec. Pap. Geol. Soc. Am.* **243**, 107–133 (1990).
- Court, N. Limb posture and gait in *Numidotherium koholense*, a primitive proboscidean from the Eocene of Algeria. *Zool. J. Linn. Soc.* **111**, 297–338 (1994).
- MacPhee, R. D. E. Morphology, adaptations, and relationships of *Plesiorcyteropus*, and a diagnosis of a new order of eutherian mammals. *Bull. Am. Mus. Nat. Hist.* **220**, 1–214 (1994).
- Godinot, M., Smith, T. & Smith, R. Mode de vie et affinités de *Paschatherium* (Condylarthra, Hyopsodontidae) d’après ses os du tarse. *Palaeovertebrata* **25**, 225–242, plate 1–2 (1996).
- Szalay, F. S. *Evolutionary History of the Marsupials and an Analysis of Osteological Characters* (Cambridge Univ. Press, Cambridge, 1994).
- Savage, R. J. G., Domning, D. P. & Thewissen, J. G. M. Fossil Sirenia of the West Atlantic and Caribbean region. V. The most primitive known sirenian, *Prorastomus sirenioides* Owen, 1855. *J. Vertebr. Paleontol.* **14**, 427–449 (1994).
- Ginsburg, L., Durrani, K. H., Kass, A. M. & Welcomme, J.-L. Discovery of a new Anthracobunidae (Tethytheria, Mammalia) from the lower Eocene lignite of the Kach-Harnai area in Baluchistan (Pakistan). *C.R. Acad. Sci. IIa* **328**, 209–213 (1999).
- Thewissen, J. G. M. Cephalic evidence for the affinities of Tubulidentata. *Mammalia* **49**, 257–284 (1985).

Supplementary Information accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We would like to thank: T. M. Bown, for collection of the original *Apheliscus* skeletal association; D. Diveley, J. Meng (American Museum of Natural History), R. Emry, L. Gordon, H. Kafka, J. Mead, R. Purdy (USNM) and P. Houde (New Mexico State University) for access to specimens; D. M. Boyer for helpful discussions; J. C. Mussell for advice and comments on the manuscript; and D. B. Weishampel for access to equipment and facilities. We would also like to thank G. B. Rathbun for access to high quality images of sengis. Support of fieldwork leading to these discoveries has been provided by the US National Science Foundation and the Bureau of Land Management.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to S.P.Z. (szack1@jhm.jhmi.edu).

## Evidence that sensory traps can evolve into honest signals

Constantino Macías García & Elvia Ramirez

Departamento de Ecología Evolutiva, Instituto de Ecología, UNAM, AP 70-275, CP 04510 México DF, México

Conventional models<sup>1–4</sup> explaining extreme sexual ornaments propose that these reflect male genetic quality<sup>2–4</sup> or are arbitrary results of genetic linkage between female preference and the ornament<sup>1</sup>. The chase-away model<sup>5</sup> emphasizes sexual conflict: male signals attract females because they exploit receiver biases<sup>6–9</sup>. As males gain control of mating decisions, females may experience fitness costs through suboptimal mating rates or post-copulatory exploitation. Elaboration of male signals is expected if females increase their response threshold to resist such exploitation. If ornaments target otherwise adaptive biases such as feeding responses<sup>8–10</sup>, selection on females might eventually separate sexual and non-sexual responses to the signal. Here we show that the terminal yellow band (TYB) of several *Goodiinae* species evokes both feeding and sexual responses; sexual



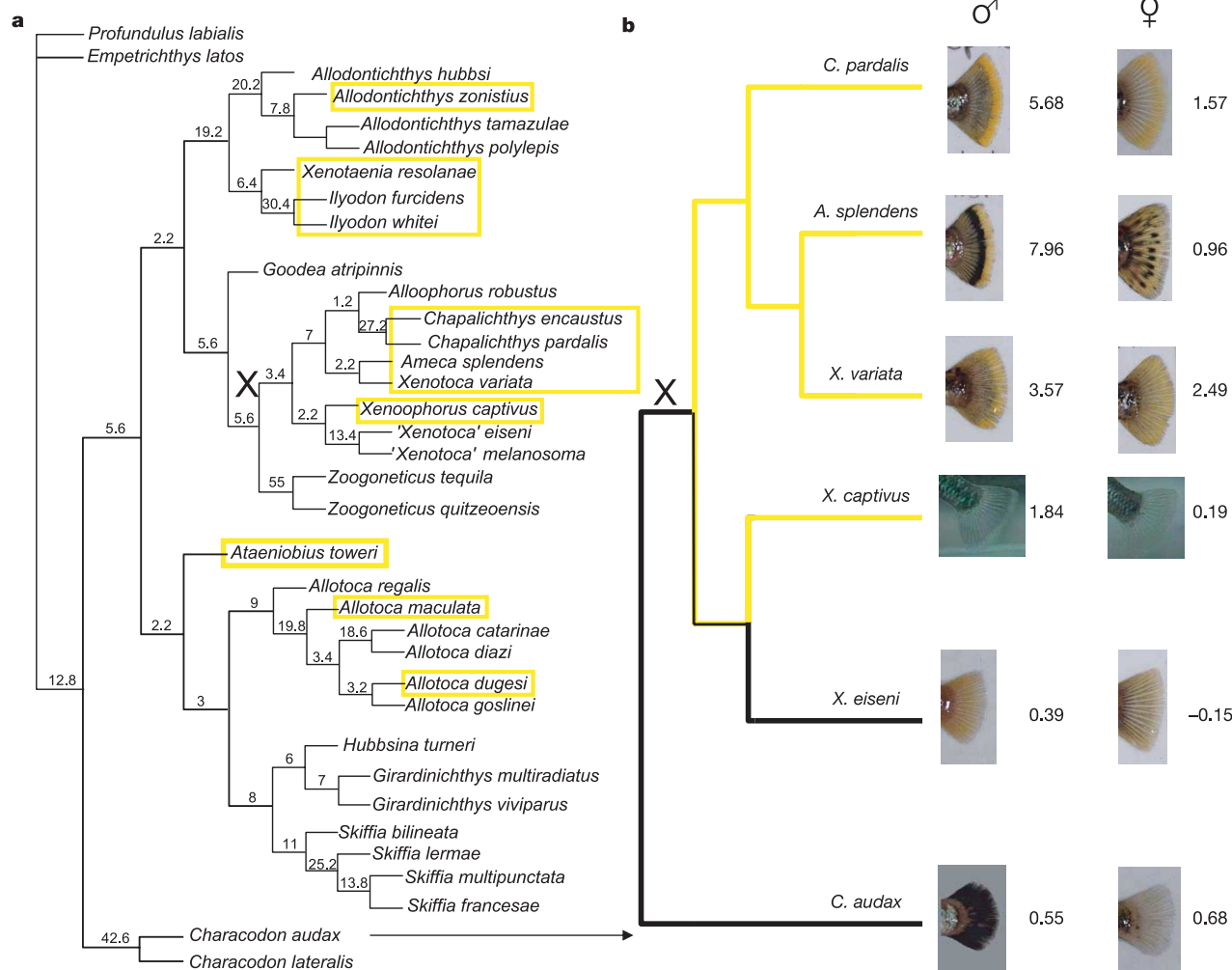
responsiveness phylogenetically pre-dates the expression of the TYB in males and is comparable across taxa, yet feeding responsiveness decreases in species with more elaborated TYBs. Displaying a TYB is costly, and thus provides an example where a trait arose as a sensory trap but has evolved into an honest signal.

We tested whether inappropriate female responses and increasing cost of male display ornaments<sup>11–13</sup> promote the separation of sexual and non-sexual responses to ornaments using a signal found in some species of the Goodeinae, a group of viviparous fish endemic to Central Mexico<sup>14</sup>. The males of several species have terminal yellow bands on their tails, which in some species are made conspicuous by the presence of a black sub-terminal band. As the tail fin undulates, the TYB resembles a quivering yellow worm or larva, with a length, movement pattern and size similar to those of damselfly larvae. The TYB occurs in only a few species, but it may have evolved independently at least six times in the group (Fig. 1). The clade containing the most species with TYBs consists of the genera *Allophorus*, *Chapalichthys*, *Ameca* and *Xenotoca* (a polyphyletic genus<sup>14</sup>). We used one species of each of the above genera, except for the monospecific *Allophorus*, a large piscivore unsuitable for use in our experiments. We also used two species of the sister group (*Xenophorus* *captivus*, with a poorly marked TYB, and

*Xenotoca* *eiseni*, whose males lack TYB but have somewhat orange fins), and one species of the most basal Goodeinae genus (*Characodon* *audax*), the males of which have black tail fins and no TYB.

We quantified the conspicuousness of the TYB from its spectral curve. We then evaluated whether females of the different species showed preferences for males with conspicuous TYBs. In experiment 1 we first evaluated whether the TYB is a sexual ornament in the species where males possess a conspicuous TYB (*C. pardalis*, *X. variata* and *A. splendens*). We then evaluated female preference for an artificial (painted) TYB in all the species; this allowed comparisons of female sexual responsiveness to males with TYBs across different species, including those lacking a TYB. In experiment 2 we quantified female and male feeding responsiveness to TYBs. We also evaluated in one species (*X. variata*) whether the lack of feeding responsiveness to the TYB of their own species could be due to resistance, by exposing the fish to increasingly conspicuous TYBs. In experiment 3 we compared the response of *C. audax* to TYBs and to damselfly larvae. In experiment 4 we evaluated whether having a conspicuous TYB is costly by measuring damage to tails of male and female *A. splendens* kept in outdoor ponds, and subsequently assessing weight loss after fin clipping.

In experiment 1 we tested whether the TYB might have evolved



**Figure 1** Tail terminal yellow band in Goodeinae. **a**, Occurrence of the TYB (yellow) in the Goodeinae phylogeny<sup>14</sup>; node support is indicated by decay indices. **b**, We used one species of each genus from the clade marked 'X' (except *Allophorus*) as well as *C. audax*. Parsimony suggests either seven independent origins of the TYB or six origins and one

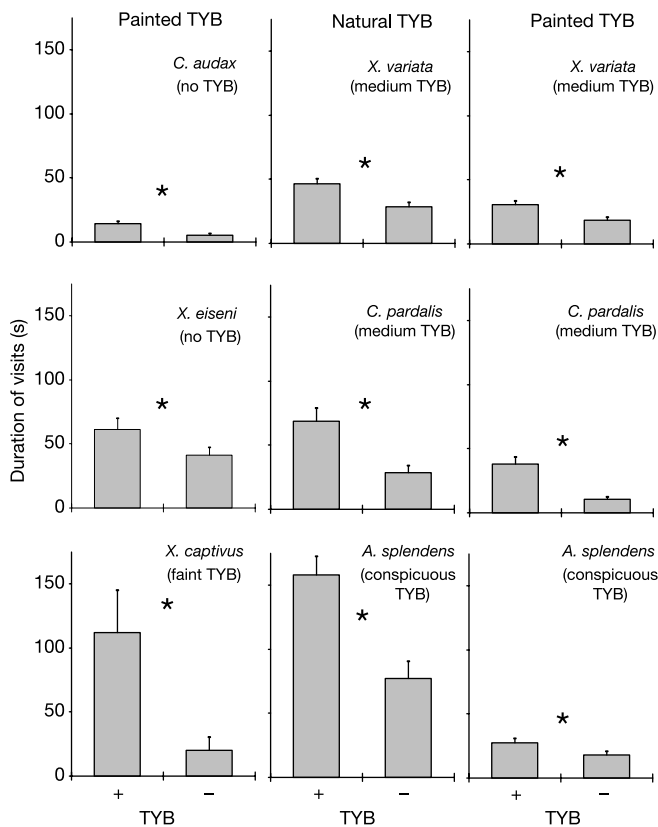
loss (bicoloured lines show ambiguous ancestral state). Values next to the images of tails are measures of TYB elaboration derived from calculating the yellow contrast of the TYB in ten males and ten females of each species (see Supplementary Information).

through female choice. Females of all species spent significantly more time visiting the male with a conspicuous TYB than visiting the male with a plainer tail (Fig. 2). In species with a conspicuous TYB, preference for males with an artificial TYB was also significant. Thus, in the species studied, the preference for males with a TYB is universal and pre-dates the expression of the male attribute (Fig. 1). In experiment 2, designed to evaluate whether the TYB evokes foraging behaviour, females and males of species lacking a conspicuous TYB bit tails with conspicuous TYBs (Fig. 3). Fish of all species visited the tail with a conspicuous TYB more often and for longer periods, although this trend was only significant for those species lacking a conspicuous TYB (Fig. 3). Fish of all species except *X. variata* and *C. pardalis* directed bites at the TYB of male *X. variata* (or of conspecific males in the case of *A. splendens*) (Fig. 3). We also found that when confronted with more conspicuous tails from heterospecific males, *X. variata* made longer and more frequent visits, and directed more bites to the tail with the conspicuous TYB ( $F_{(3,35)} = 3.03$ ,  $P = 0.04$ ; see Supplementary Information). In experiment 3 we evaluated whether the tail edge of a species with a conspicuous TYB acts as a stronger feeding stimulus than a damselfly larva. Faced with a choice of a larva and a tail with a TYB, *C. audax* approached and attacked first the tail with the TYB ( $\chi^2_1 = 7.2$ , where subscript denotes d.f.;  $P = 0.0073$ ), and only afterwards turned to the larva. In experiment 4 we verified that tails of male and female *A. splendens* were of similar size (tail length

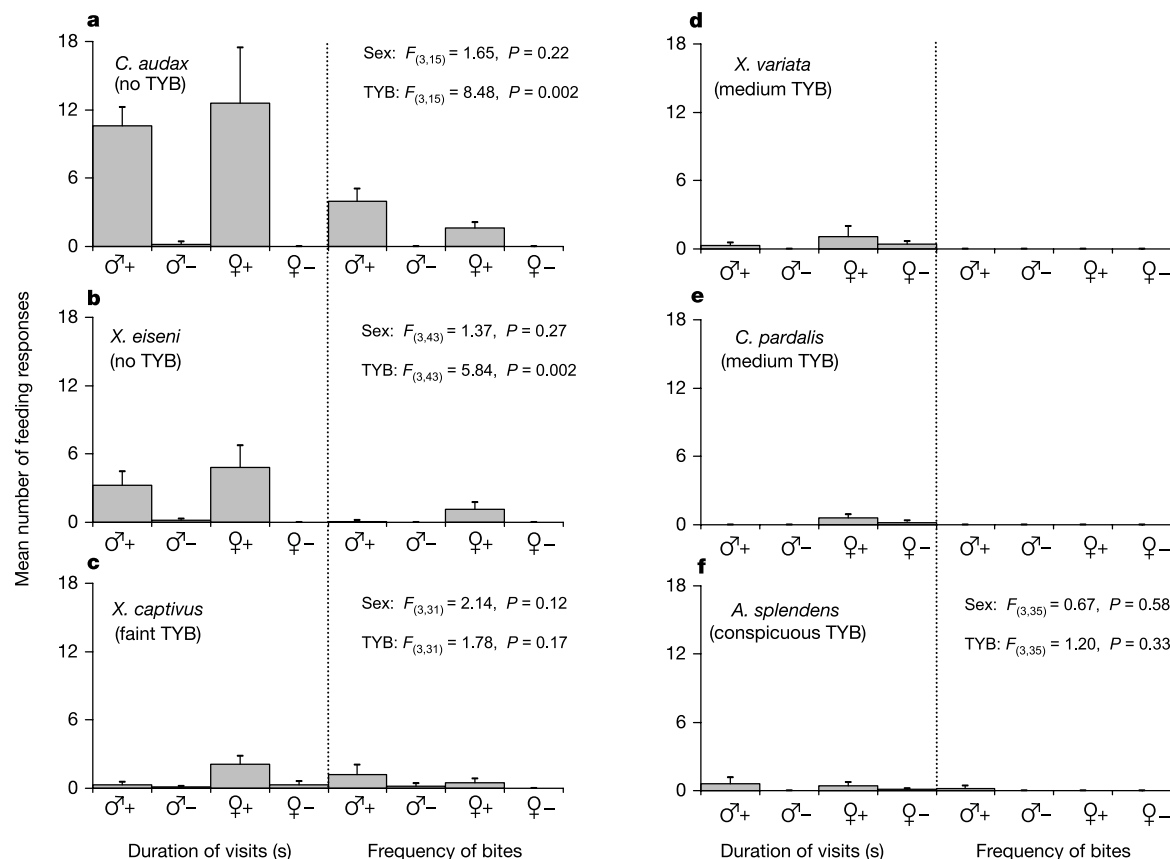
given as mean  $\pm$  s.d.; male,  $7.8 \pm 0.4$  mm; female,  $7.4 \pm 1.3$  mm;  $t_{11} = -0.85$ ,  $P = 0.41$ ), and that the area of the measured band was also similar (mean  $\pm$  s.d.; male,  $44.3 \pm 5.6$  mm<sup>2</sup>; female,  $40.6 \pm 14.1$  mm<sup>2</sup>;  $t_{13} = -0.80$ ,  $P = 0.44$ ). Male tails were significantly more damaged than female tails; the damage was largely confined to the area of the TYB (mean damaged area  $\pm$  s.d.; male,  $5.1 \pm 2.5$  mm<sup>2</sup>; female,  $1.6 \pm 1.2$  mm<sup>2</sup>;  $t_{14} = -4.06$ ,  $P = 0.0012$ ). Males subjected to fin clipping lost weight in proportion to the amount of tissue regenerated, as evidenced by the negative association between the rate of weight change and the proportion of tissue regenerated ( $r^2 = 0.34$ ,  $F_{(1,10)} = 5.04$ ,  $P = 0.049$ ; see Supplementary Information).

The conspicuous terminal yellow band of some Goodeinae evokes feeding responses in both males and females. Regardless of whether they mimic the shape and movement pattern of damselfly larvae or of another prey item, elaborated TYBs are powerful feeding stimuli. They probably evolved through exaggeration of prey attributes in order to retain female interest, given that conspicuous TYBs provide stronger stimuli than readily consumed larvae, and that fish of species with conspicuous yellow bands show reduced feeding responsiveness to TYBs than fish from species with poorly developed or absent TYBs. The decrease in female feeding responsiveness with increasing (conspecific) male ornament elaboration (Fig. 4), whether through an increase in response threshold (that is, resistance<sup>5</sup>) or because females evolve the ability to discriminate between the model (food) and the mimic (TYB), does not lead to a decrease in female responsiveness to males displaying a TYB. Instead, feeding disappears but the TYB still functions to attract females (Fig. 4). We suggest that selection has led females to disentangle the feeding and sexual responses, so that they remain responsive to a frequent prey item that would otherwise provide an insufficient stimulus if females merely increased their resistance to the signal produced by the mimic<sup>5</sup>. This solution may also permit females to incorporate a new criterion for mate choice<sup>15</sup>: the expression of an honest ornament in the form of a TYB. We did not quantify indirect fitness benefits for the females, but have shown that displaying a TYB entails costs in terms of damaging bites and regeneration expenditure. Another potential cost stems from the brightness of TYBs<sup>4,16</sup>, which may require diverting carotenoids from other functions<sup>17</sup>.

Both the ornament and the responsiveness to it could, in principle, be present in both sexes<sup>10,18,19</sup>. In fact, sex-limitation of the TYB is not universal, as both male and female *Chapalichthys* spp. have TYBs. Females in species with rudimentary TYBs and in species with the most conspicuous ornament (Fig. 1) lack a TYB. We suggest that this pattern arises because selection for males with TYBs in species with incipient markings (a condition probably represented by *X. captivus*) leads to conspicuous yellow bands in both sexes (as in *Chapalichthys* spp., Fig. 1). As the cost of displaying the TYB mounts, this promotes mechanisms that limit its expression in the females (such as *A. splendens*). A similar process may have led to the testosterone-determined control of the expression of orange spots in the trinidadian guppy (*Poecilia reticulata*)<sup>20</sup>, and might be a mechanism of sex-limitation that contributes to the transformation of a sensory trap into a condition-dependent ornament<sup>21</sup>. On the other hand, feeding responsiveness to the TYB, when present, is similar in males and in females (Fig. 3). This may imply that males and females have roughly the same diet<sup>10</sup> (if the visual system and related neural structures are not sexually dimorphic, which need not be the case<sup>22</sup>). Female and male feeding responsiveness to TYBs decreased with TYB elaboration (Fig. 3). This may mean that both sexes face similar costs of responding to a food mimic. It is possible that such reduction in responsiveness is due to increased resistance to avoid manipulation (in females)<sup>23,24</sup> or to avoid a maladaptive response (in males). One alternative is that the means to differentiate between food and the TYB evolved in both males and females. This would permit both

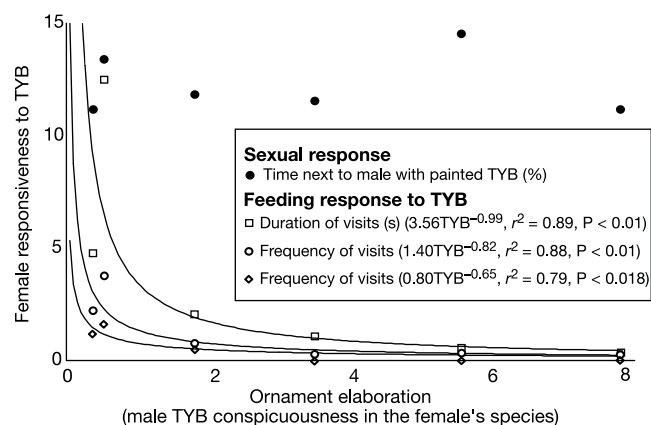


**Figure 2** Female preference for males with elaborate TYBs. Preference for conspecific males with conspicuous (+) or inconspicuous (−) TYBs was assessed using either natural variation (central column; *X. variata*, *C. pardalis* and *A. splendens*) or artificial (painted) TYBs. Mean time spent next to each male during the 5-min trials ( $\pm$  s.e.m.) was contrasted using a balanced analysis of variance (ANOVA) design with three factors (treatment, pair of males, and female) for each species. Here we show only the results for the first factor (treatment: intensity of TYB; see Supplementary Information). Asterisks denote significant differences ( $P < 0.05$ ).



**Figure 3** Feeding responsiveness to tails with (+) and without a TYB (–). Different pairs of tail-exhibiting *X. variata* (a–d) or *A. splendens* (e, f) were used in each trial, which involved the simultaneous observation of two focal fish. Differences in responsiveness (mean  $\pm$  s.e.m.) between TYB treatments (+ and –) and between sexes were

simultaneously compared using a two-way multiple analysis of variance (MANOVA) for each species. Dependent variables were frequency of bites, frequency of visits (not shown) and duration of visits (see Supplementary Information). Tests were not possible with *X. variata* and *C. pardalis* owing to an excess of ‘zero’ scores.



**Figure 4** Female responsiveness and male ornament elaboration. Feeding responsiveness (open symbols) decreases significantly as ornament elaboration increases (male TYB conspicuousness in the female's own species; Fig. 1), whereas sexual responsiveness (to standard, artificial TYB; closed circles) remains significant and substantial across species. Instead of using the tails of *X. variata*, the TYB signal presumed to provide the strongest stimulus (*A. splendens*' TYB) was conservatively used to test feeding responses of species with conspicuous TYBs (*C. pardalis* and *A. splendens*). Strong sexual responsiveness across taxa suggests that ornament elaboration in clades with conspicuous TYBs retains a high level of female interest that cannot be explained by feeding responsiveness.

sexes to escape a sensory trap without compromising foraging efficiency. We recorded few feeding responses by male *A. splendens* to a TYB when presented next to a plain tail, but pilot trials showed that this species frequently bite the TYB in the presence of a damselfly larva, suggesting that additional stimuli are required to trigger feeding behaviour in species with conspicuous TYBs. Although we have no evidence for the mechanism involved, it seems that females in some species of this fish clade escaped a sensory trap<sup>9</sup> while retaining sexual responsiveness to a male attribute that has become a costly signal. □

## Methods

### TYB conspicuousness

Ten males and ten females of each species were kept in outdoor ponds either on their own (*Ameca splendens*) or with other species (*X. eiseni*, *C. audax*, *X. captivus*, *C. pardalis* and *X. variata*; see Supplementary Information); a further ten male and ten female *C. audax* were collected shortly before trials. Fish were anaesthetized (using 20 ml of a 2% dilution of benzocaine in acetone per 100 ml of water) and laid on a white board. The reflectance spectra of the terminal and sub-terminal bands were measured using a Minolta CM-2600D spectrophotometer with pulsed xenon lamps (360–740 nm). Spectra were used to calculate an index of TYB elaboration (see Supplementary Information). Recovery was accelerated using aeration and fin conditioner (Stress-Coat, Aquarium Pharmaceuticals).

### Female sexual responsiveness

Fish were housed individually between trials. Responsiveness to natural TYB variation was investigated in *C. pardalis* ( $n = 9$ ), *X. variata* and *A. splendens* ( $n = 10$  each). These were exposed to 4 (*C. pardalis*) or 5 pairs of males of equal size but with contrasting TYB intensity (+ or –) as judged by two independent observers. Experimental tanks ( $50 \times 25 \times 25$  cm deep aquaria) were divided into one female ( $25 \times 25 \times 25$  cm) and two male compartments ( $12.5 \times 25 \times 25$  cm, separated by mirrors to standardize the male social environment). Males of one pair were placed in their compartments, and each



female was exposed to them in turn. A female was placed in the centre of the female compartment within a plastic bag for 5 min, then released and the number and duration of visits (approaches within one body length) to each male was recorded for 5 min (a common index of mate choice<sup>25,26</sup>, which correlates with mating probability in the goodeid *Girardinichthys multiradiatus*). She was then transferred to her home tank and a new female was tested until all had been exposed to that male pair. This was repeated on consecutive days until all females had been exposed to all male pairs. Position (left or right) of the TYB + male changed between pairs. Female preference in *X. captivus* (faint TYB) and *X. eiseni* and *C. audax* (no TYB) was assessed by painting an artificial TYB to one male of each pair using nail varnish. We proceeded as above but after completing the trials with a male pair, the treatment was reversed and all females were tested again. We used 10 females and 4 male pairs. This protocol was repeated, painting the tails of different species with conspicuous TYB (*A. splendens*, *C. pardalis* and *X. variata*).

### Feeding responsiveness

Fish were kept without food for 24 h in the presentation tank, then exposed to two fish kept in green plastic boxes (4 × 2.5 × 2.5 cm) with adjustable walls that enabled only their tails to protrude. Boxes were hung at opposite sides of the tank, and gently rocked with a motor to stimulate tail movement. Before each trial, a partition hid the presentation boxes from the focal fish. Stimuli fish were one male with a conspicuous TYB and one female *X. variata* lacking a TYB (focal *C. pardalis* and *A. splendens* were exposed to *A. splendens* as stimuli). To minimize stress, stimuli fish were briefly exposed to a weak dilution of benzocaine and placed in the presentation boxes. The motor was then started and the partition was removed. Frequency and duration of approaches and frequency of bites to each tail were recorded for 10 min, then both focal and stimuli fish were replaced by new pairs, and the procedure was repeated until all fish had been tested. We also exposed *X. variata* to stimuli fish of species with more conspicuous TYBs (*A. splendens* and *C. pardalis*). Feeding responsiveness of *C. audax* (10 males and 10 females, tested in groups) to TYBs and damselfly larvae was assessed with a similar procedure, except that the male *X. variata* was presented alongside a damselfly larva (*Argia* sp.) of the same length as the male's TYB. We recorded which stimulus was approached first and which was bitten first.

### Cost of the TYB

Digital images of tail fins of 11 male and 11 female *A. splendens* from outdoor ponds were measured using Paint Shop Pro v. 7. We traced TYB area, both bridging the gaps ('intact' band), and following the contour of tissue present (actual band). The difference measured the amount of tissue lost. Females lack a TYB but have a discontinuous black sub-terminal band seemingly homologous to the black sub-terminal band that limits the male TYB (Fig. 1). This was used in tracing the area of the female terminal band. The effect of fin regeneration on weight loss was evaluated by removing sections ( $23 \pm 7.4\%$ ) of fin tail in 12 anaesthetized males. These were kept in individual 40-litre aquaria under controlled conditions (including equal feeding opportunity and antibiotics), and were measured, photographed and weighed weekly for two months.

Received 29 November 2004; accepted 14 January 2005; doi:10.1038/nature03363.

1. Fisher, R. A. *The Genetical Theory of Natural Selection* 146–156 (Dover Publications, New York, 1958).
2. Andersson, M. Evolution of condition-dependent sex ornaments and mating preferences: sexual selection based on viability differences. *Evolution* **40**, 804–816 (1986).
3. Hamilton, W. D. & Zuk, M. Heritable true fitness and bright birds: a role for parasites? *Science* **218**, 384–387 (1982).
4. Milinski, M. & Bakker, T. C. M. Female sticklebacks use male coloration in mate choice and hence avoid parasitized males. *Nature* **344**, 330–333 (1990).
5. Holland, B. & Rice, W. R. Chase-away sexual selection: antagonistic seduction versus resistance. *Evolution* **52**, 1–7 (1998).
6. Ryan, M. J. Sexual selection, sensory systems and sensory exploitation. *Oxf. Surv. Evol. Biol.* **7**, 157–195 (1990).
7. Ryan, M. J. & Keddy-Hector, A. Directional patterns of female mate choice and the role of sensory biases. *Am. Nat.* **139**, S4–S35 (1992).
8. Dawkins, M. S. & Guilford, T. Sensory bias and the adaptiveness of female choice. *Am. Nat.* **148**, 937–942 (1996).
9. Endler, J. A. & Basolo, A. L. Sensory ecology, receiver biases and sexual selection. *Trends Ecol. Evol.* **13**, 415–420 (1998).
10. Rodd, F. H., Hughes, K. A., Grether, G. F. & Baril, C. T. A possible non-sexual origin of mate preference: are male guppies mimicking fruit? *Proc. R. Soc. Lond. B* **269**, 475–481 (2002).
11. Kokko, H., Brooks, R., McNamara, J. M. & Houston, A. I. The sexual selection continuum. *Proc. R. Soc. Lond. B* **269**, 1331–1340 (2002).
12. Kokko, H., Brooks, R., Jennions, M. D. & Morley, J. The evolution of mate choice and mating biases. *Proc. R. Soc. Lond. B* **270**, 653–664 (2003).
13. Cordero, C. & Eberhard, W. G. Female choice of sexually antagonistic male adaptations: a critical review of some current research. *J. Evol. Biol.* **16**, 1–6 (2003).
14. Webb, S. A. *et al.* Molecular phylogeny of the live-bearing Goodeidae (Cyprinodontiformes). *Mol. Phylogenet. Evol.* **30**, 527–544 (2004).
15. LeBas, N., Hockham, L. R. & Ritchie, M. G. Nonlinear and correlational sexual selection on 'honest' female ornamentation. *Proc. R. Soc. Lond. B* **270**, 2159–2165 (2003).
16. Kodric-Brown, A. Dietary carotenoids and male mating success in the guppy: an environmental component to female choice. *Behav. Ecol. Sociobiol.* **25**, 393–401 (1989).
17. Olson, V. A. & Owens, I. P. F. Costly sexual signals: are carotenoids rare, risky or required? *Trends Ecol. Evol.* **13**, 510–514 (1998).
18. Partridge, L. & Parker, G. A. in *Evolution of Biological Diversity* (eds Magurran, A. E. & May, R. M.) 130–159 (Oxford Univ. Press, Oxford, 1999).
19. Basolo, A. L. Congruence between the sexes in preexisting receiver responses. *Behav. Ecol.* **13**(6), 832–837 (2002).

20. Jayasooriya, A. P., Weisinger, R. S., Weisinger, H. S., Mathai, M. L. & Sinclair, A. J. Attraction to orange: sexiness, not gluttony. *Science* **296**, 847–848 (2002).
21. Grether, G. F., Hughes, K. A. & Rodd, F. H. Response to Jayasooriya *et al.* *Science* **296**, 847–848 (2002).
22. Burton, B. G. & Laughlin, S. B. Neural images of pursuit targets in the photoreceptor arrays of male and female houseflies *Musca domestica*. *J. Exp. Biol.* **206**, 3963–3977 (2003).
23. Córdoba-Aguilar, A. Sensory trap as the mechanism of sexual selection in a damselfly genital trait (insecta: calopterygidae). *Am. Nat.* **160**, 594–601 (2002).
24. Holland, B. & Rice, W. R. Experimental removal of sexual selection reverses intersexual antagonistic coevolution and removes a reproductive load. *Proc. Natl Acad. Sci. USA* **96**, 5083–5088 (1999).
25. Ryan, M. J. & Wagner, E. Asymmetries in mating preferences between species: female swordtails prefer heterospecific males. *Science* **236**, 595–597 (1987).
26. Basolo, A. L. Evolutionary change in a receiver bias: a comparison of female preference functions. *Proc. R. Soc. Lond. B* **265**, 2223–2228 (1998).

Supplementary Information accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We Thank the Lunáticos behaviour discussion group for constructive discussion throughout, and M. Ritchie, J. Graves and H. Kokko for useful comments on previous versions of the manuscript. A. Gaona and M. Molina helped with some of the behavioural records, and E. Ávila Luna provided logistical aid with fish husbandry and data collection. This project was partly supported with a grant from CONACYT to C.M.G. and is based on E.R.'s BSc thesis (Biology), which was supervised by C.M.G.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to C.M.G. ([maciasg@servidor.unam.mx](mailto:maciasg@servidor.unam.mx)).

## Genome-wide non-mendelian inheritance of extra-genomic information in *Arabidopsis*

Susan J. Lolle\*, Jennifer L. Victor, Jessica M. Young & Robert E. Pruitt\*

Department of Botany and Plant Pathology, Purdue University, 915 W. State Street, West Lafayette, Indiana 47907-2054, USA

\* These authors contributed equally to this work

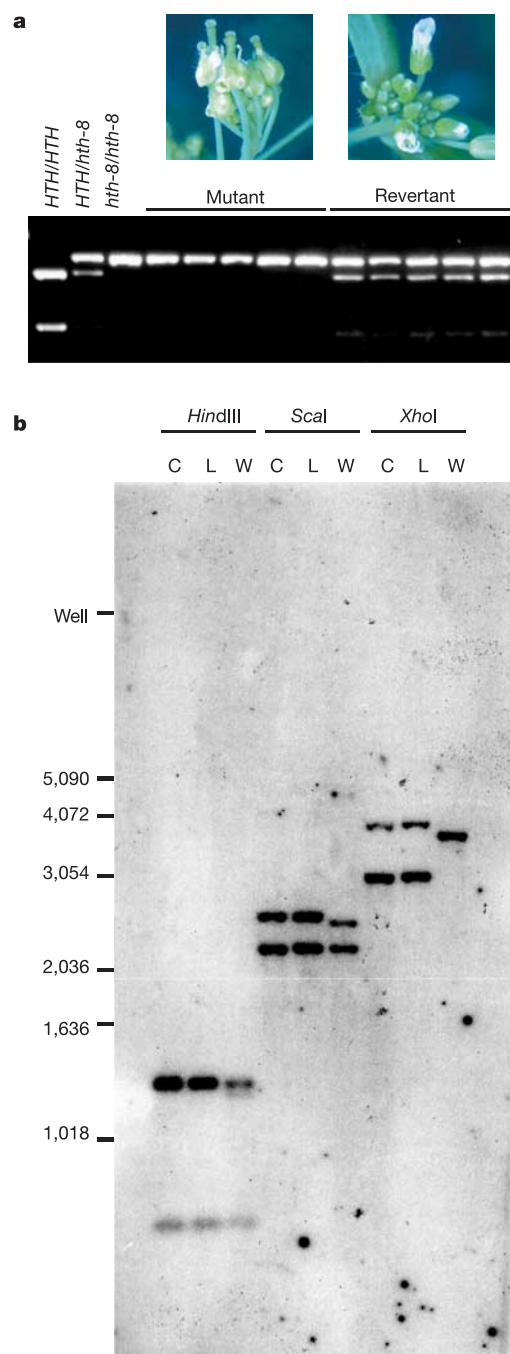
A fundamental tenet of classical mendelian genetics is that allelic information is stably inherited from one generation to the next, resulting in predictable segregation patterns of differing alleles<sup>1</sup>. Although several exceptions to this principle are known, all represent specialized cases that are mechanistically restricted to either a limited set of specific genes (for example mating type conversion in yeast<sup>2</sup>) or specific types of alleles (for example alleles containing transposons<sup>3</sup> or repeated sequences<sup>4</sup>). Here we show that *Arabidopsis* plants homozygous for recessive mutant alleles of the organ fusion gene *HOTHEAD*<sup>5</sup> (*HTH*) can inherit allele-specific DNA sequence information that was not present in the chromosomal genome of their parents but was present in previous generations. This previously undescribed process is shown to occur at all DNA sequence polymorphisms examined and therefore seems to be a general mechanism for extra-genomic inheritance of DNA sequence information. We postulate that these genetic restoration events are the result of a template-directed process that makes use of an ancestral RNA-sequence cache.

We have recovered 11 point mutations at the *hth* locus<sup>6</sup> that share the unusual property that they segregate phenotypically wild-type plants at a high frequency when homozygous mutant plants are allowed to self-fertilize. The frequency with which these 'revertant' plants are recovered varies, but is generally in the range of  $10^{-1}$  to  $10^{-2}$  revertants per chromosome per generation (Table 1). This is in stark contrast to most point mutations, which are completely stable (for example *erecta* (*er*) in Table 1). Because of the high frequency

with which *hth* phenotypic revertants were observed, we suspected that they might reflect either incomplete penetrance of the mutant phenotype or an epigenetic change that masked the mutant genotype. To determine the genotype of the phenotypic revertants we designed allele-specific polymerase chain reaction (PCR)-based markers. These molecular markers show clearly that the phenotypically wild-type plants observed are heterozygous for the parental *hth* allele (Fig. 1a, Table 1), indicating that the nucleotide sequence of the *hth* gene had been altered in about 10% of the progeny plants. This was confirmed by allowing the revertant plants to self-fertilize and examining the segregation of the *hth* phenotype in the next generation. In every case the mutant phenotype was observed in about 25% of the progeny plants (data not shown).

Two trivial explanations could account for the observed revertant progeny: contamination of our stocks with wild-type seed or outcrossing of the *hth* mutant plants with wild-type pollen. Seed contamination was an unlikely explanation because the apparent revertants are all heterozygous for the specific *hth* allele carried by their parents. To rule out all possibility of seed contamination, individual embryos were dissected out of fruits developing on selfed plants that were genotypically confirmed to be *hth/hth* by using allele-specific PCR-based markers. The genotypes of these embryos with respect to *HTH* were then determined with the same markers. The results in Table 2 show that we can detect frequent *HTH/hth* embryos developing on *hth/hth* parent plants, eliminating the possibility that the observed high frequency of reversion is due to seed contamination. In fact, in this experiment we also detected rare double revertant *HTH/HTH* embryos, which must have inherited one of their two wild-type *HTH* genes from the maternal parent and therefore could not have been the result of outcrossing. The frequency of apparent reversion events observed in embryos is somewhat higher than that observed phenotypically. We believe this probably reflects somatic reversion events occurring in embryonic tissues that do not contribute to the adult plant.

To demonstrate directly that *hth/hth* mutant plants can be a source of wild-type *HTH* alleles, we tested whether it was possible to transmit an *HTH* allele through pollen derived from an *hth/hth* plant. Plants homozygous for *hth-4* in the Landsberg *erecta* (Ler) background were reciprocally crossed with wild-type Columbia (Col) plants and the embryos arising from the crosses were genotyped with respect to *HTH*. Most embryos were heterozygous as expected, but in the crosses where the male parent was homozygous for *hth-4*, 8 of 164 embryos were observed that were of the genotype *HTH/HTH* (reversion frequency  $4.9 \times 10^{-2}$ ). Because the Col and Ler accessions differ at numerous DNA sequence polymorphisms, we used accession-specific DNA markers to verify that the *HTH/HTH* embryos were not due to self-pollination of the Col parent. No *HTH/HTH* embryos were detected when the female parent was homozygous for *hth-4* (0 of 230 embryos). These data indicate that *hth/hth* plants can be a source of pollen bearing an *HTH* allele and furthermore that there is a strong bias for genetic changes to occur in the male reproductive system. Taken together, the results show that the wild-type *HTH* alleles observed



**Figure 1** Molecular genetic analysis of *HTH*. **a**, PCR-based assay demonstrating genetic reversion of *hth-8*. DNA was extracted from mutant and revertant progeny that were derived from the self-fertilization of a homozygous *hth-8* parent, amplified with gene-specific primers and digested to reveal the presence or absence of a polymorphic restriction site. Controls in the left three lanes show digest patterns derived from wild-type (*HTH/HTH*), heterozygous (*HTH/hth*) and homozygous (*hth/hth*) mutant plants. Photos above mutant and revertant lanes show the *hth/hth* mutant phenotype and the revertant phenotype. **b**, Genome blot showing that *HTH* is present in a single copy in Col (C), Ler (L) and Ws (W). Genomic DNA was digested as indicated and probed with a labelled fragment of Col DNA corresponding to the *HTH* coding sequence. Expected fragment sizes in base pairs are as follows: for *HindIII*: Col 1,285, 1,253, 664; Ler 1,275, 1,253, 664; Ws 1,253, 1,204, 666; for *SalI*: Col 2,537, 2,187; Ler 2,527, 2,187; Ws 2,466, 2,188; for *XhoI*: Col 3,928, 3,016; Ler 3,918, 3,016; Ws 3,669. Differences in hybridization intensity occur because portions of the Ws sequence have diverged significantly from the Col sequence.

**Table 1** Phenotypic and genotypic reversion of *hth* alleles

| Mutation      | Phenotype*  |           | Genotype of revertants† |                |                | Reversion frequency‡ |
|---------------|-------------|-----------|-------------------------|----------------|----------------|----------------------|
|               | Mutant      | Revertant | <i>hth/hth</i>          | <i>hth/HTH</i> | <i>HTH/HTH</i> |                      |
| <i>hth-4</i>  | 491 (92)    | 43 (8)    | 0 (0)                   | 38 (100)       | 0 (0)          | $4.0 \times 10^{-2}$ |
| <i>hth-8</i>  | 184 (84)    | 36 (16)   | 0 (0)                   | 24 (100)       | 0 (0)          | $8.2 \times 10^{-2}$ |
| <i>hth-10</i> | 251 (88)    | 35 (12)   | 0 (0)                   | 30 (100)       | 0 (0)          | $6.1 \times 10^{-2}$ |
| <i>er</i>     | 1,000 (100) | 0 (0)     | —                       | —              | —              | 0                    |

Numbers in parentheses are the percentage of individuals with a given phenotype or genotype. \*Observed numbers of phenotypically mutant and wild-type progeny derived from the self-fertilization of plants homozygous for the mutant alleles shown in the first column. †Genotype of phenotypically wild-type plants determined by PCR analysis. Not all revertant individuals were genotyped. ‡Reversion frequency (reversion events per chromosome per generation) calculated on the basis of the observed number of phenotypic revertants and the observed genotypes of the revertants.

in progeny of selfed *hth/hth* mutant plants are not the result of seed contamination or cross-pollination but represent true genetic events manifested at the DNA sequence level.

We considered several conventional explanations for the genetic instability observed at *hth*. DNA sequence analysis of the *HTH* gene rules out the involvement of transposons and repeated sequences as possible explanations. Two additional explanations that we considered were a high rate of random mutation in this particular region of the genome and correction of the gene through a gene conversion mechanism.

If reversion of *hth* to *HTH* were due to a high rate of random mutation at this locus, we would expect to find additional silent sequence changes in the reverted *HTH* allele. Given the frequency with which the mutant nucleotide is observed to revert to wild-type and the 1,226 possible silent nucleotide substitutions in the *HTH* coding sequence, we would expect to see between 49 and 100 other changes on average in the coding sequence of the reverted alleles, assuming that all types of nucleotide substitutions are equally likely. To test this we determined the complete DNA sequence of the coding region of the *HTH* gene after three independent reversion events in each of three independently derived mutant *hth* alleles (*hth-4*, *hth-8* and *hth-10*). In every case the sequence of the reverted *HTH* allele matched the Ler wild-type sequence exactly. Given the very high frequency with which the mutant nucleotide reverts to the wild-type sequence, this result rules out a mechanism dependent on random sequence changes and strongly indicates a mechanism that is template-directed.

In considering the possibility of gene conversion we needed to exclude two types of source template. First, it was possible that the *Arabidopsis* Ler and Wassilewskija (Ws) accessions contained a second copy of the *HTH* gene that was not present in the Col genome sequence. This is unlikely because we have designed many PCR-based genetic markers that allow genotyping of mutant *hth* alleles and none revealed a cryptic wild-type copy of the gene in homozygous *hth* plants. To examine this possibility further, we performed a DNA blotting experiment with genomic Col, Ler and Ws DNA. We were able to detect all of the expected *HTH* fragments in this experiment, and no additional hybridizing sequences were identified (Fig. 1b). Taken together, the PCR and genomic hybridization results effectively rule out the existence of a second cryptic copy of the *HTH* gene.

The second class of template we considered was the family of *HOTHEAD-LIKE* (*HTL*) genes<sup>6</sup>. For a variety of reasons this possibility could also be discounted. For *hth-4*, none of the *HTL* sequences contained the appropriate nucleotide to correct the mutant sequence to wild-type (Fig. 2). For *hth-8* and *hth-10*, some of the *HTL* sequences did contain the relevant nucleotide to correct the mutant sequence, but if any of these related sequences had been used as a corrective template this would have resulted in the incorporation of other nearby DNA sequence changes, unless the conversion tract were limited to just a few nucleotides (Fig. 2). Because no other DNA sequence changes were observed, it seems unlikely that conventional gene conversion is the explanation for the high level of genetic change seen at the *HTH* locus.

In the experiments described above we examined the process of

DNA sequence change at the *hth* locus in an *hth* mutant background. To determine whether DNA sequence changes also occur elsewhere in the genome, we crossed *hth-4*, *hth-8* and *hth-10* mutant plants (in the Ler genetic background) to wild-type Col plants. The *F*<sub>1</sub> hybrids were allowed to self-fertilize and individual *F*<sub>2</sub> plants were genotyped with the use of PCR-based markers to identify those that were homozygous for the mutant or wild-type allele of *HTH*. These *F*<sub>2</sub> individuals were then genotyped for several molecular markers that are polymorphic between Col and Ler, to identify individuals that were homozygous for one parental allele or the other. Randomly chosen *F*<sub>3</sub> progeny from these plants were in turn examined for the stable inheritance of those DNA sequence polymorphisms. In progeny of *hth/hth* *F*<sub>2</sub> plants we observed a high rate of sequence change in both directions (Col to Ler, and Ler to Col), indicating that the mechanism of genetic instability observed at *hth* seems to apply globally to other sequences in the genome (Table 3). Although the polymorphisms examined are all within genes, their locations are in exons, introns and the 3' untranslated region of those genes (Table 3), indicating that presence in the mature mRNA is not required for extra-genomic transmission of sequence information. We also detected phenotypic reversion of the *er* mutation in the progeny of *hth/hth er/er* *F*<sub>2</sub> plants at similar frequencies, demonstrating that the absence of *HTH* leads to the destabilization of *er*, which is in contrast to the stability seen in an *HTH/HTH* background (Table 1). Similarly, inheritance of molecular markers in the progeny of the *HTH/HTH* *F*<sub>2</sub> plants was also completely stable, indicating that the instability observed is dependent on the absence of *HTH* gene function.

Mutation of the *Arabidopsis* *HTH* gene reveals an unusual pattern of genetic transmission in which progeny plants inherit, at a relatively high frequency, DNA sequences different from those

```

HTH      ATTCGGCCGTGTCACACCGC
hth-4     ATTCGGCCGTGTCACACCGC
HTL1      caaCGGCaaTaGACACACCGC
HTL2      gTgtGGgaaaGacATACatC
HTL3      gTgtGGgaaaGacATACatC
HTL4      tactGGgaGgaGACACAGCtC
HTL5      ggatGGtCaTaGACACAGGC
HTL6      tgctGGtCaTaGACATACCGC
HTL7      tgctGGtCaTaGACACAGGC

```

```

HTH      CGAGTCTCCAGGAACCAACCC
hth-8     CGAGTCTCCAGGAACCAACCC
HTL1      ttAtTgtCCCGgACaAAtCC
HTL2      aGAaTcaCCAGGAACaAACCC
HTL3      aGAaTcaCCAGGAACaAACCC
HTL4      tatcTCTCCAGgActAAtCC
HTL5      CaAGTCTCCGGgActAAtCC
HTL6      CaAGTCTCCCGgActAAtCC
HTL7      CaAGTCTCCGGgActAAtCC

```

```

HTH      CAGACTGTTGGAATTACAAAG
hth-10    CAGACTGTTGGAATTACAAAG
HTL1      gAGgtTGTGGgATTACggg
HTL2      gAtcCaccTcaAgTTgtAgcG
HTL3      gAGcCaccTcaAgTTgCAGca
HTL4      CAagtCGTgCGgtTtActgAG
HTL5      CAagCcGTaGGAATcACcAAG
HTL6      CAagCTGTGGgATcACAAAG
HTL7      CAagtCGTTGGAATcACAAAG

```

**Figure 2** DNA sequences of *HTH* and *HTL* genes around the sites of mutation in *hth-4*, *hth-8* and *hth-10*. The location of the mutant nucleotide is boxed, and the wild-type nucleotide at this position is shown in blue; differences from the *HTH* sequence are shown in red lower-case letters. For *hth-4* none of the related sequences contain the wild-type nucleotide at the site of the mutation, for *hth-8* all of the other sequences contain the appropriate wild-type nucleotide, and for *hth-10* some of the other sequences contain the relevant wild-type nucleotide. In all cases there are sequence differences between the *HTL* sequences and *HTH* within a few nucleotides of the site of the mutation.

**Table 2** Genotypic reversion of embryonic progeny of *hth/hth* parents

| Mutation      | Genotype of embryos* |                |                | Reversion frequency† |
|---------------|----------------------|----------------|----------------|----------------------|
|               | <i>hth/hth</i>       | <i>hth/HTH</i> | <i>HTH/HTH</i> |                      |
| <i>hth-4</i>  | 174 (75)             | 57 (24)        | 2 (1)          | $1.3 \times 10^{-1}$ |
| <i>hth-8</i>  | 67 (62)              | 41 (38)        | 0 (0)          | $1.9 \times 10^{-1}$ |
| <i>hth-10</i> | 222 (84)             | 43 (16)        | 0 (0)          | $8.1 \times 10^{-2}$ |

Numbers in parentheses are the percentage of embryos with a given genotype.

\* Observed numbers of embryos with each genotype derived from the self-fertilization of plants homozygous for the mutant alleles shown in the first column.

† Reversion frequency (reversion events per chromosome per generation) calculated on the basis of the observed frequencies of embryonic genotypes.



carried by their parents. Although this was initially observed as a result of the instability of nucleotide substitution mutations in the *hth* gene itself, we have subsequently shown that loss of *HTH* gene function also results in the genetic instability of DNA sequences located throughout the *Arabidopsis* genome. However, the instability observed is not random but seems to be confined to the restoration of sequence information in progeny plants that, although absent from the parent genome, was present in the genome of an earlier ancestor. Thus, we believe that mutant *hth* alleles are able to revert to wild type because recent ancestors of the *hth/hth* mutant plants had a wild-type copy of the *HTH* gene and, similarly, that we can identify sequence polymorphism changes in plants whose parents were homozygous for one form of the polymorphism because they had grandparents that were heterozygous for the same polymorphism. Sequencing of multiple independently reverted *hth* alleles revealed no DNA sequence changes other than the restoration of the mutant nucleotide to the wild type, also indicating that the process of sequence change is non-random.

Similar patterns of inheritance have been observed previously under various exceptional circumstances. It has been reported that molecular markers absent from several generations of a population of synthetic polyploid plants can reappear subsequently in the population<sup>7</sup>. It has also been reported that random amplified polymorphic DNA (RAPD) markers can in rare instances be transmitted from stock to scion after grafting<sup>8</sup>. In addition to reports of aberrant inheritance of molecular markers, there are also numerous reports of 'replicating instabilities', which seem to propagate both wild-type and mutant alleles in a single copy of a gene<sup>9</sup>. All of these occurred at much lower frequencies than we have observed in the *hth* mutant background, but could be explained by a common mechanism involving the non-mendelian inheritance of sequence information maintained outside the conventional DNA genome.

On the basis of the high degree of specificity seen in the observed DNA sequence changes, we propose that this is a template-directed process. Because we have been unable to detect a DNA template by either genomic DNA blotting or PCR, a logical alternative is that the template is instead provided by an RNA molecule. We propose a model in which a type of stable RNA, possibly in a double-stranded form, can be replicated and transmitted over multiple generations. This template must under some circumstances be capable of modifying the DNA sequence of the nuclear genome to restore sequence information cached from previous generations. Given that this type of sequence modification occurs globally throughout the genome, we think it likely that the postulated genome-wide sequence cache also exists in wild-type plants and that the *hth* mutation simply increases the frequency with which DNA sequences are modified.

Although the model proposed above might represent an extraordinary view of inheritance, each of the processes required by the model has been demonstrated previously. Research performed over the past few years in post-transcriptional gene silencing or RNA interference has shown that silencing induced by double-stranded

RNA can persist for several generations in *Caenorhabditis elegans*<sup>10</sup> and can spread through the organism in plants, fungi and animals<sup>11,12</sup>. This strongly implies that some form of nucleotide sequence information must be stably replicated and transmitted outside the DNA genome, which is consistent with a requirement for an RNA-directed RNA polymerase in post-transcriptional gene silencing<sup>13–15</sup>. Replication of the sequence cache could allow it to persist for several generations; this is consistent with the fact that *hth* mutations retain their ability to revert even after multiple generations in the homozygous state (S.J.L. and R.E.P., unpublished observations). Experiments to detect changes in the reversion frequency after several generations of homozygous propagation are in progress. Both double-stranded RNA and microRNA have been implicated in the sequence-specific modification of the nuclear genome by DNA methylation<sup>16–18</sup> and it has been possible to perform site-directed mutagenesis by the introduction of chimaeric RNA–DNA oligonucleotides in both plant and animal cells<sup>19–21</sup>. We propose that a mechanism combining these individual steps could explain the exceptional inheritance we see in *hth* mutants as well as the unusual results observed by earlier workers.

It is tempting to speculate on the function of this unusual mechanism of inheritance in the wild-type organism. On the basis of our observation that progeny of *hth* mutant plants seem to inherit a wide variety of sequence information missing from the DNA genomes of their parents, one possibility is that this mechanism provides a cache of genetic variation beyond that carried in the chromosomes. If this information can then be copied back into the DNA genome it could allow self-fertilizing species such as *Arabidopsis* to avoid the negative consequences of inbreeding. Perhaps even more intriguing is the possibility that the inheritance of non-genomic templates occurs in wild-type and *hth* plants, but the rate at which those templates are used to modify genomic sequences is elevated in *hth* as a result of an indirect 'stress' put on the plant by the absence of the HTH gene product. Under these circumstances one could envisage a mechanism in which additional allelic information is maintained outside the normal genomic context but could be used under conditions that compromised the continued functioning of the organism. The extra-genomic information used by the organism would not be random in this case but would contain a library of potential allelic sequences that were previously carried in the DNA of its ancestors, all of which must have survived to reproduce. These sequences would therefore represent a collection of alleles already known to be functional under at least some circumstances, and the restoration of these sequences to the DNA genome might represent a mechanism for the organism to recover from a genotype that was poorly adapted to its present environment. Any sequence changes that resulted in the elimination of the 'stress' would also decrease the rate of sequence alteration back to its normal background level, effectively fixing a favourable genotype. This is at least superficially similar to some models proposed to account for 'adaptive' mutation in micro-organisms<sup>22–24</sup> and it is possible that recent work describing increased rates of genomic change in response to environmental stress<sup>25,26</sup> and pathogen attack<sup>27</sup> reflect the action of such a mechanism. It is also possible that some of the cryptic genetic variation uncovered through the loss of Hsp90 function<sup>28,29</sup> is actually present in the postulated ancestral sequence cache and is restored to the DNA genome in response to a lower level of Hsp90 activity.

Whatever molecular mechanism leads to the unusual pattern of genetic inheritance we have observed in *hth* mutants, the genetic results presented here indicate the existence of an extra-genomic mechanism that leads to the high-frequency modification of DNA sequences in a template-directed manner. The structure of the templates together with the mechanisms by which they are produced, inherited and used remain questions for future research. It will be of great interest to see how widely this mechanism of inheritance is distributed throughout the tree of life. □

**Table 3 Genetic instability of molecular markers in the *hth* background**

| Marker | F <sub>2</sub> genotype (parents) |                | Location in gene |
|--------|-----------------------------------|----------------|------------------|
|        | <i>hth/hth</i>                    | <i>HTH/HTH</i> |                  |
| AG     | 7/186 (4)                         | 0/190 (0)      | Intron           |
| GAPC   | 9/242 (4)                         | 0/190 (0)      | Intron           |
| GL1    | 1/90 (1)                          | 0/196 (0)      | 3'-UTR           |
| HTH    | 10/484 (2)                        | 0/590 (0)      | Exon             |
| RGA    | 14/402 (3)                        | 0/386 (0)      | Exon             |
| UFO    | 16/438 (4)                        | 0/196 (0)      | Exon             |

Fractions in the table represent the observed number of non-parental alleles/total number of chromosomes examined for a non-selected sample of F<sub>3</sub> progeny derived by the self-fertilization of F<sub>2</sub> plants with the genotype indicated at the top of the table. Numbers in parentheses indicate the percentage of non-parental alleles observed in the F<sub>3</sub> progeny. UTR, untranslated region. Further information on markers is available on TAIR (<http://www.arabidopsis.org>).

## Methods

DNA extraction, blotting, labelling and hybridization techniques were as described previously<sup>5</sup>. Genomic DNA sequences were determined by amplifying portions of the *hth* gene and directly sequencing the PCR product<sup>30</sup>. DNA samples for the PCR genotyping of embryos were obtained by dissecting individual embryos from siliques and surrounding maternal seed coat tissue, crushing the embryo with a plastic pestle in a Microfuge tube and resuspending the resulting homogenate in 20 µl of TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). A 1 µl portion of this sample was used in a standard PCR reaction.

Received 8 December 2004; accepted 21 January 2005; doi:10.1038/nature03380.

- Mendel, G. Versuche über Pflanzen Hybriden. *Verhandl. Naturforsch. Ver. Brünn* **4**, 3–47 (1866).
- Klar, A. J., Fogel, S. & Lusnak, K. Gene conversion of the mating-type locus in *Saccharomyces cerevisiae*. *Genetics* **92**, 777–782 (1979).
- McClintock, B. The origin and behavior of mutable loci in maize. *Proc. Natl Acad. Sci. USA* **36**, 344–355 (1950).
- Gondo, Y. *et al.* High-frequency genetic reversion mediated by a DNA duplication: the mouse pink-eyed unstable mutation. *Proc. Natl Acad. Sci. USA* **90**, 297–301 (1993).
- Lolle, S. J., Hsu, W. & Pruitt, R. E. Genetic analysis of organ fusion in *Arabidopsis thaliana*. *Genetics* **149**, 607–619 (1998).
- Krolkowski, K. A., Victor, J. L., Nussbaum Wagler, T., Lolle, S. J. & Pruitt, R. E. Isolation and characterization of the *Arabidopsis* organ fusion gene *HOTHEAD*. *Plant J.* **35**, 501–511 (2003).
- Song, K., Lu, P., Tang, K. & Osborn, T. C. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proc. Natl Acad. Sci. USA* **92**, 7719–7723 (1995).
- Taller, J., Hirata, Y., Yagishita, N., Kita, M. & Ogata, S. Graft-induced genetic changes and the inheritance of several characteristics in pepper (*Capsicum annuum* L.). *Theor. Appl. Genet.* **97**, 705–713 (1998).
- Auerbach, C. & Kilbey, B. J. Mutation in eukaryotes. *Annu. Rev. Genet.* **5**, 163–218 (1971).
- Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
- Cogoni, C. *et al.* Transgene silencing of the *al-1* gene in vegetative cells of *Neurospora* is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation. *EMBO J.* **15**, 3153–3163 (1996).
- Palauqui, J. C., Elmayan, T., Pollien, J. M. & Vaucheret, H. Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. *EMBO J.* **16**, 4738–4745 (1997).
- Cogoni, C. & Macino, G. Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature* **399**, 166–169 (1999).
- Dalmay, T., Hamilton, A., Rudd, S., Angell, S. & Baulcombe, D. C. An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* **101**, 543–553 (2000).
- Mourrain, P. *et al.* *Arabidopsis* SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell* **101**, 533–542 (2000).
- Wassenegger, M., Heimes, S., Riedel, L. & Sanger, H. L. RNA-directed de novo methylation of genomic sequences in plants. *Cell* **76**, 567–576 (1994).
- Mette, M. F., Aufsatz, W., van der Winden, J., Matzke, M. A. & Matzke, A. J. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.* **19**, 5194–5201 (2000).
- Bao, N., Lye, K.-W. & Barton, M. K. MicroRNA binding sites in *Arabidopsis* class III HD-ZIP mRNAs are required for methylation of the template chromosome. *Dev. Cell* **7**, 653–662 (2004).
- Cole-Strauss, A. *et al.* Correction of the mutation responsible for sickle cell anemia by an RNA-DNA oligonucleotide. *Science* **273**, 1386–1389 (1996).
- Beetham, P. R., Kipp, P. B., Sawycky, X. L., Arntzen, C. J. & May, G. D. A tool for functional plant genomics: chimeric RNA/DNA oligonucleotides cause *in vivo* gene-specific mutations. *Proc. Natl Acad. Sci. USA* **96**, 8774–8778 (1999).
- Zhu, T. *et al.* Targeted manipulation of maize genes *in vivo* using chimeric RNA/DNA oligonucleotides. *Proc. Natl Acad. Sci. USA* **96**, 8768–8773 (1999).
- Hall, B. G. Spontaneous point mutations that occur more often when advantageous than when neutral. *Genetics* **126**, 5–16 (1990).
- Foster, P. L. & Cairns, J. Mechanisms of directed mutation. *Genetics* **131**, 783–789 (1992).
- Galitski, T. & Roth, J. R. A search for a general phenomenon of adaptive mutability. *Genetics* **143**, 645–659 (1996).
- Ries, G. *et al.* Elevated UV-B radiation reduces genome stability in plants. *Nature* **406**, 98–101 (2000).
- Bjedov, I. *et al.* Stress-induced mutagenesis in bacteria. *Science* **300**, 1404–1409 (2003).
- Kovalchuk, I. *et al.* Pathogen-induced systemic plant signal triggers DNA rearrangements. *Nature* **423**, 760–762 (2003).
- Queitsch, C., Sangster, T. A. & Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624 (2002).
- Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
- Pruitt, R. E., Vielle-Calzada, J. P., Ploense, S. E., Grossniklaus, U. & Lolle, S. J. *FIDDLEHEAD*, a gene required to suppress epidermal cell interactions in *Arabidopsis*, encodes a putative lipid biosynthetic enzyme. *Proc. Natl Acad. Sci. USA* **97**, 1311–1316 (2000).

**Acknowledgements** We thank R. M. Lee and T. Nussbaum Wagler for technical assistance, and J. A. Banks, J. L. Bowman, U. Grossniklaus, R. A. Jorgensen, J. Ogas, D. R. Smyth, A. Steimer, V. Sundaresan and M. A. Webb for discussions. This work was supported by the National Science Foundation. Any opinion, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to R.E.P. (pruitt@purdue.edu).

# Independent recruitment of a conserved developmental mechanism during leaf evolution

C. Jill Harrison<sup>\*</sup>, Susie B. Corley<sup>\*</sup>, Elizabeth C. Moylan<sup>†</sup>, Debbie L. Alexander<sup>†</sup>, Robert W. Scotland & Jane A. Langdale

Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK

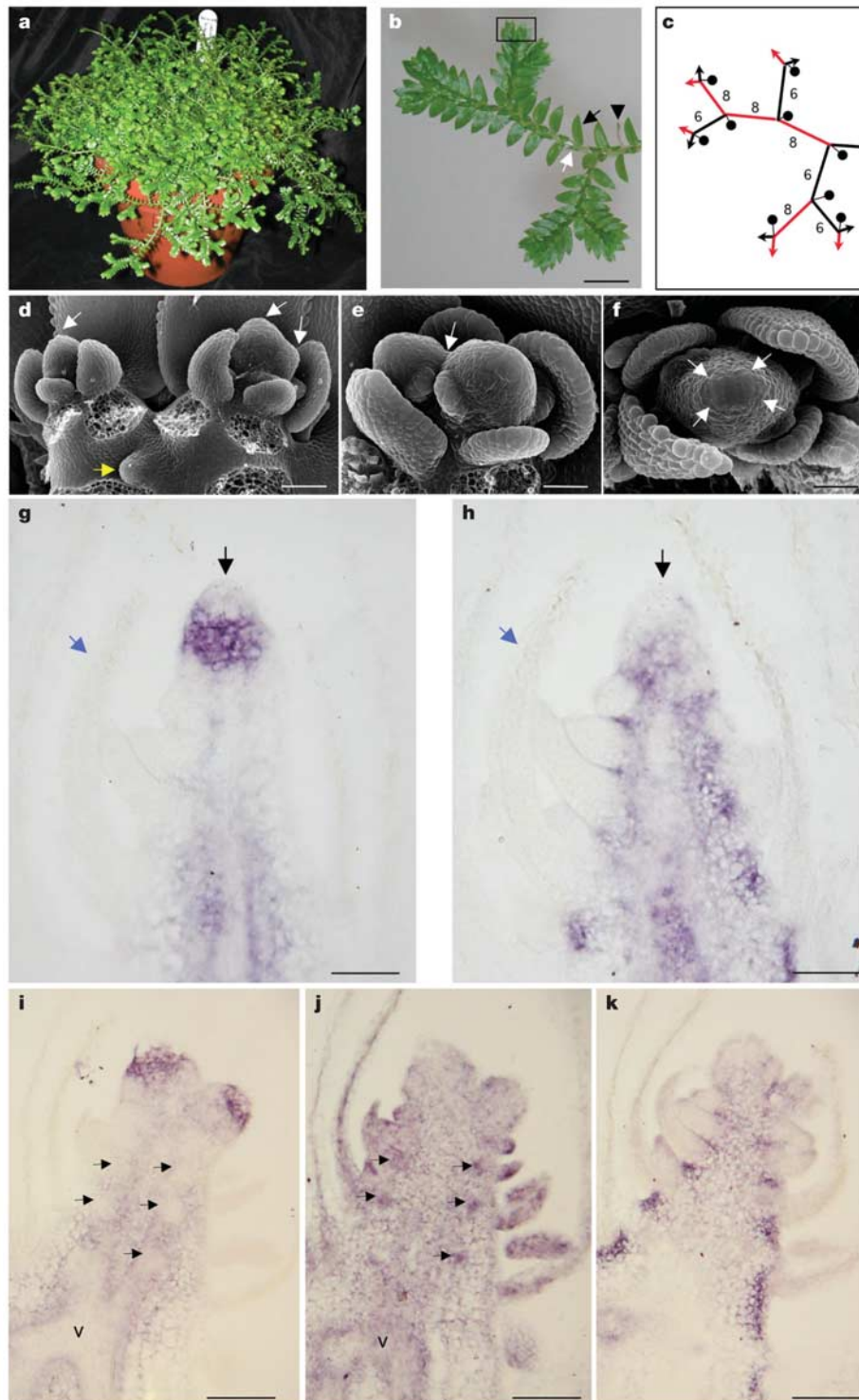
<sup>\*</sup> These authors contributed equally to this work

<sup>†</sup> Present addresses: BioMed Central, Middlesex House, 32–42 Cleveland Street, London W1T 4LB, UK (E.C.M.); Carnegie Institution of Washington, Department of Plant Biology, 260 Panama Street, Stanford, California 94305, USA (D.L.A.)

Vascular plants evolved in the Middle to Late Silurian period, about 420 million years ago<sup>1</sup>. The fossil record indicates that these primitive plants had branched stems with sporangia but no leaves. Leaf-like lateral outgrowths subsequently evolved on at least two independent occasions<sup>2–4</sup>. In extant plants, these events are represented by microphyllous leaves in lycophytes (club-mosses, spikemosses and quillworts) and megaphyllous leaves in euphyllophytes (ferns, gymnosperms and angiosperms). Our current understanding of how leaves develop is restricted to processes that operate during megaphyll formation. Because microphylls and megaphylls evolved independently, different mechanisms might be required for leaf formation. Here we show that this is not so. Gene expression data from a microphyllous lycophyte, phylogenetic analyses, and a cross-species complementation experiment all show that a common developmental mechanism can underpin both microphyll and megaphyll formation. We propose that this mechanism might have operated originally in the context of primitive plant apices to facilitate bifurcation. Recruitment of this pathway to form leaves occurred independently and in parallel in different plant lineages.

Microphylls and megaphylls are determinate organs produced on the flanks of indeterminate shoot apical meristems (SAMs)<sup>2</sup>. The formation of all vascular plant leaves therefore involves the addition of a determinate growth programme to the indeterminate apical growth programme. Microphylls develop simply with a single vascular trace, whereas megaphylls develop complex vasculature and variable shape. The genetic basis of the developmental transition from indeterminate growth in the apex to determinate growth in the leaf has so far been studied only in euphyllophyte species with megaphyllous leaves. Indeterminate apical growth is marked by class I *knotted1*-like homeobox (*KNOX*) gene expression<sup>5–12</sup>. Conversely, determinate leaf growth is marked by transcriptional<sup>5–10</sup>, or possibly post-transcriptional<sup>12</sup>, repression of *KNOX* activity. In *Arabidopsis*, maize and *Antirrhinum*, MYB orthologues (*ASYMMETRIC LEAVES1*, *ROUGH SHEATH2* and *PHANTASTICA* (*ARP*), respectively) maintain the *KNOX*-off state in leaves<sup>13–16</sup>. *KNOX* and *ARP* genes are expressed in mutually exclusive domains. In loss-of-function *arp* and gain-of-function *KNOX* mutants, ectopic foliar *KNOX* expression leads to indeterminate growth such that simple leaves become lobed<sup>6,13–17</sup>. Therefore, in these three species and also in tobacco<sup>18</sup>, *KNOX*–*ARP* interactions in the shoot apex regulate the balance between indeterminate and determinate growth.

*KNOX*–*ARP* interactions facilitate megaphyll formation in both monocots and eudicots, suggesting that aspects of leaf development are conserved between groups that diverged about 140 million years ago<sup>19</sup>. Involvement of the same mechanism in the formation of microphylls would imply the independent recruitment of identical processes in species that diverged more than 350 million years ago. To investigate the meristem-to-leaf transition in a lycophyte, we have examined *KNOX*–*ARP* interactions in *Selaginella kraussiana*, a



**Figure 1** KNOX-ARP relationships in *S. kraussiana*. **a**, Mature plant. **b**, Shoot showing position of large lower leaves (black arrow), small upper leaves (white arrow) and rhizophores (black arrowhead). Scale bar, 0.5 cm. **c**, Diagram of branching pattern seen in **b**; red lines depict major branches, black lines depict minor branches and circles depict rhizophores. Numbers indicate the number of leaf pairs on each branch. **d**, Scanning electron micrograph (SEM) of region outlined in **b**, showing shoot meristems (white arrows) and an emerging rhizophore (yellow arrow). Scale bar, 25  $\mu$ m. **e**, SEM of branching shoot. Arrow points to bifurcating meristem. Scale bar, 9  $\mu$ m. **f**, SEM of shoot apex showing strip of large cells at the apex (flanked by white arrows). Scale bar, 8  $\mu$ m.

**g, h**, *In situ* hybridization of *SkKNOX1* (**g**) and *SkKNOX2* (**h**) in sequential sagittal sections of a *Selaginella* shoot apex. Black arrows point to the apical cells; blue arrows point to leaf primordia. In each case, the entire coding region 5' to the ELK domain was used as a hybridization probe. Scale bar, 38  $\mu$ m. **i-k**, Hybridization of *SkKNOX1* (**i**), *SkARP1* (**j**) and *SkKNOX2* (**k**) to sequential frontal sections of a bifurcating *Selaginella* apex. Black arrows point to the insertion point of the leaf vascular trace in the stem, and v marks the stem vasculature. The *SkARP1* full-length cDNA was used as a hybridization probe. Scale bar, 85  $\mu$ m.

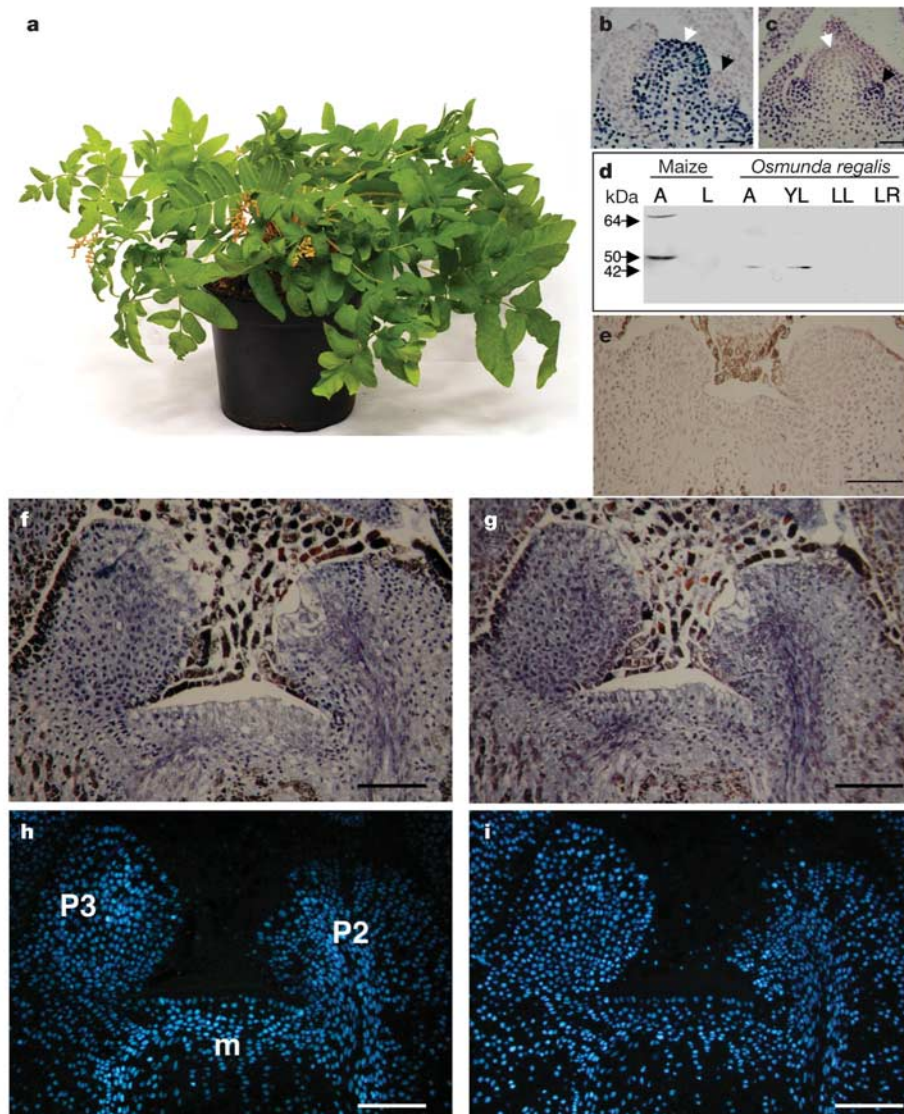


diploid lycophyte (Fig. 1a). Shoots have a dorsi-ventral organization, and leaf primordia arise in pairs, with one primordium giving rise to a large ventral (lower) leaf and the other to a small dorsal (upper) leaf. At maturity, leaves are ranked on the shoot with small leaves on the upper surface and large leaves on the lower side or in a lateral position (Fig. 1b). Shoots branch regularly after the formation of six or eight leaf pairs, and branch points are marked by the growth of a leaf pair and an aerial root-like structure known as a rhizophore on the upper surface (Fig. 1b–d). Branching involves bifurcation of the SAM (Fig. 1d, e) as opposed to outgrowth of axillary meristems as in angiosperms. A distinguishing feature of *S. kraussiana* meristems is the presence of a strip of large cells on the surface of the apex (Fig. 1f). How these cells relate to specific zones or layers of seed plant meristems is not yet clear.

*KNOX* and *ARP* gene copy number was assessed in *S. kraussiana* after gene isolation. Sequencing and hybridization analyses of

genomic DNA revealed two class I *KNOX* genes (*SkKNOX1* and *SkKNOX2*), one class II *KNOX* gene (*SkKNOX3*) and a single *ARP* gene (*SkARP1*) (Supplementary Figs S1 and S2a, c). A single *ARP* gene was also isolated from a related species, *S. viticulosa* (Supplementary Fig. S2c). Phylogenetic analyses of 112 *KNOX* genes and 125 *MYB* genes demonstrated monophyly of the class I and class II *KNOX* clades (Supplementary Figs S2b and S3), and of the *ARP* clade (Supplementary Figs S2d and S4) respectively.

Expression patterns of *SkKNOX* and *SkARP* genes were assessed by *in situ* hybridization. *SkKNOX1* was expressed in cells subtending the large superficial apical cells (Fig. 1g, i), whereas *SkKNOX2* was expressed in internodal regions (Fig. 1h, k). These patterns strikingly resemble those seen in seed plant apices<sup>20</sup>. Significantly, *SkKNOX* transcripts were not detected in leaf primordia. In contrast, *SkARP1* was expressed in leaf primordia and in the meristem (Fig. 1j). *SkARP1* expression in leaves is consistent with a mutually



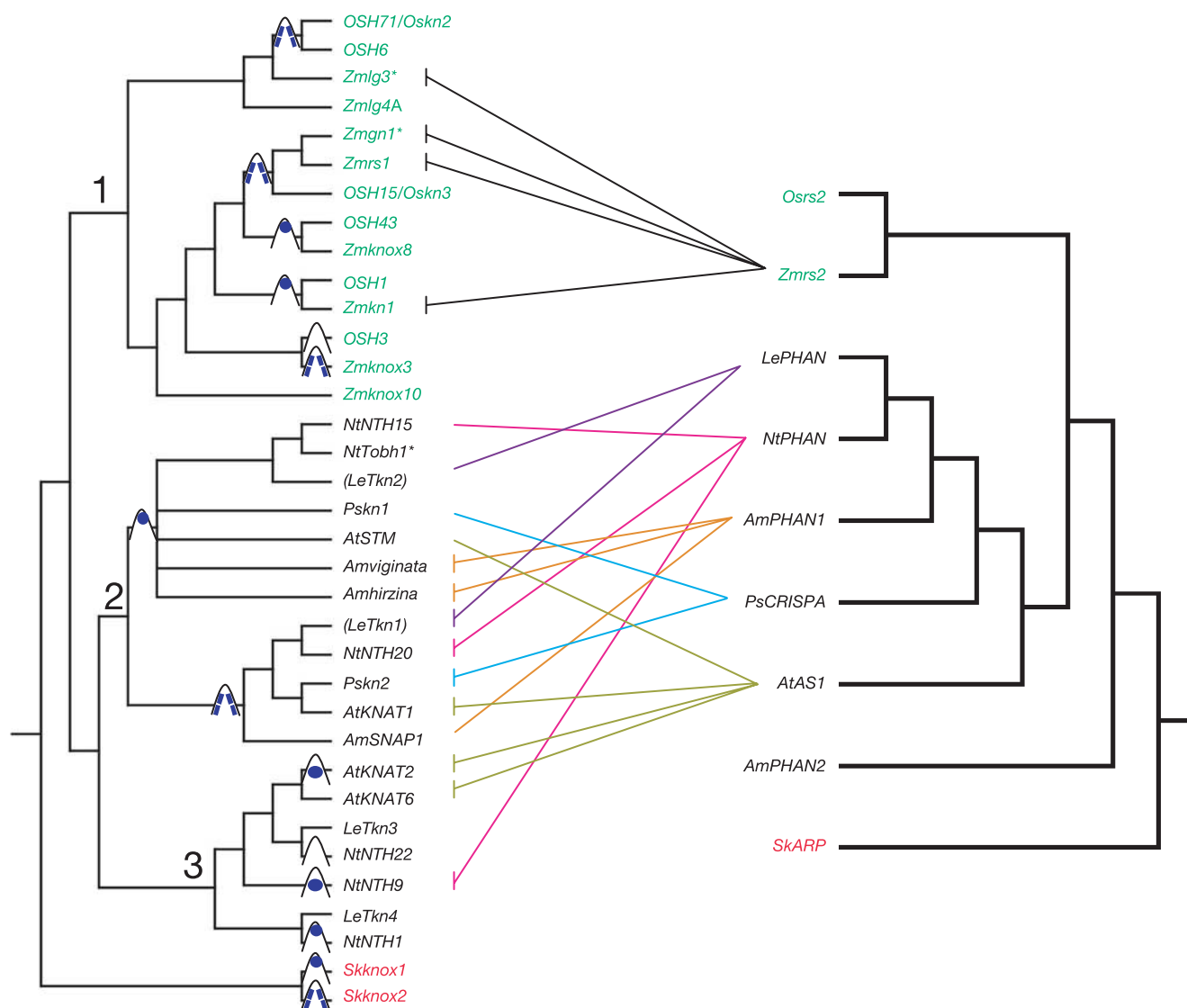
**Figure 2** KNOX–ARP relationships in *O. regalis*. **a**, Mature plant. **b**, **c**, Immunolocalization of KNOX (**b**) and ARP (**c**) proteins in maize apices. White arrows indicate the meristem and black arrows leaf primordia. Scale bar, 40  $\mu$ m. **d**, Immunoblot of proteins extracted from maize apex (A) and leaf (L) and from fern apex (A), young leaf (YL), mature leaf lamina (LL) and mature leaf rachis (LR) with anti-KNOX antibody. Proteins are detected in the maize apex but not the maize leaf. In *Osmunda*, proteins are detected in the apex

and young leaf but not in mature leaf tissues. **e**, Control immunolocalization performed with no primary antibody. Scale bar, 170  $\mu$ m. **f**, **g**, Immunolocalization of KNOX (**f**) and ARP (**g**) proteins in serial sagittal sections of *Osmunda* apices. **h**, **i**, the same sections as in **f** and **g** showing nuclei stained with 4,6-diamidino-2-phenylindole. Meristems (m) and plastochron 2 and 3 (P2 and P3) leaves are labelled. Scale bar, 180  $\mu$ m.

exclusive *KNOX*–*ARP* relationship in microphylls. Mutually exclusive expression was also seen in parts of the stem, where *ARP* but not *KNOX* genes are expressed in the stem vasculature and in the leaf traces (Fig. 1i, j). In the meristem, however, expression domains of *SkKNOX* and *SkARP1* overlap. This overlap might facilitate repression of *SkKNOX* at the midline of the apex during bifurcation.

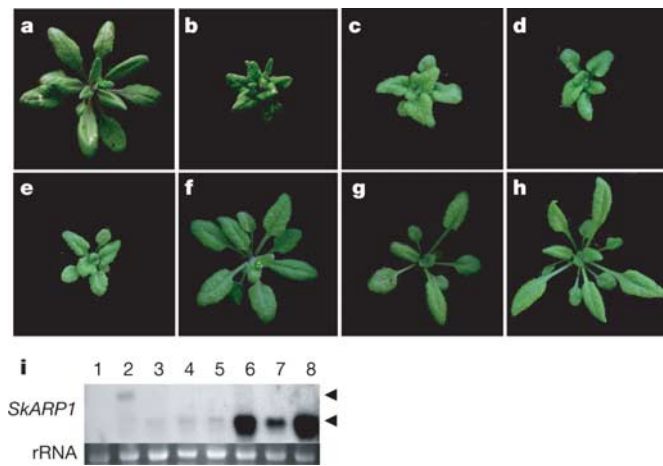
To determine whether co-expression of *KNOX* and *ARP* genes is specific to *S. kraussiana* meristems or is more broadly representative of bifurcating shoots, we examined protein localization patterns in *Osmunda regalis*. *O. regalis* is a basal leptosporangiate fern with a clump-forming growth habit (Fig. 2a). Leaves arise spirally, and in the closely related species *O. cinnamomea* they are not determined until plastochron 8–10 (P8–10)<sup>21</sup>. As in *Selaginella*, the meristem

has distinct apical cells on the surface, and bifurcates during branching<sup>22</sup>. In control experiments, *KNOX* and *ARP* proteins were detected in meristems and leaves of maize, respectively (Fig. 2b, c). In contrast, both proteins were detected in meristems of *Osmunda* (Fig. 2e–i). Co-localization of *KNOX* and *ARP* proteins was also seen in *Osmunda* leaf primordia (Fig. 2f, g). *KNOX* protein accumulation in fern leaves has previously been reported in *Anogramma chaeophylla*, where no *KNOX* repression at P0 was seen<sup>23</sup>. It was argued that this lack of repression indicated the recruitment of independent mechanisms of leaf formation in ferns and seed plants<sup>23</sup>. We postulate instead that failure to repress *KNOX* expression at P0 reflects the delayed determinacy of fern leaves<sup>21</sup>. Our immunoblot analysis supports this suggestion, because *KNOX*



**Figure 3** *KNOX*–*ARP* relationships across land plants. Phylogenetic trees of class I *KNOX* (left) and *ARP* (right) genes showing *KNOX* expression patterns and *KNOX*–*ARP* interactions. *Selaginella* (red), monocot (green) and eudicot (black) sequences are placed. The trees are redrawn from more densely sampled trees shown in Supplementary Fig. S2. *KNOX* expression patterns fall into four classes, represented diagrammatically: expression in the central zone (blue circle), in the peripheral zone (blue oblongs), in the rib zone (blue oval) or absent from the meristem (empty). Asterisks indicate cases in which gene expression patterns have not been determined, and parentheses indicate genes that are also expressed in leaf tissue. Lines between the trees indicate the relationship between *ARP* and *KNOX* genes in particular species. When loss of *ARP* function leads to changes in

*KNOX* gene expression, the line linking the two genes terminates in a vertical line. When *KNOX* gene expression is not altered by loss of *ARP* function the genes are linked by a straight line. *KNOX* expression patterns have been reported as follows: *OSH1*, *OSH3*, *OSH6*, *OSH15*, *OSH43*, *OSH71* (ref. 9); *rs1*, *kn1*, *knox3*, *knox8* (ref. 5); *KNAT1* (ref. 6); *STM*; *KNAT2* (ref. 27); *Pskn1* (ref. 10); *Pskn2* (J. Hofer, personal communication); *NTH1*, *NTH9*, *NTH15*, *NTH20*, *NTH22* (ref. 7); *SNAP1* (A. Hudson, personal communication); *invaginata*, *hirzina*<sup>28</sup>, *Tkn1* (ref. 11); *Tkn2* (refs 29, 30). The effects of loss of *ARP* function have been reported as follows: *Zmrs2* (ref. 14); *AmPHAN1* (A. Hudson, personal communication<sup>15,16</sup>); *AtAS1* (ref. 13); *NIPHAN* (N. McHale, personal communication<sup>18</sup>); *LePHAN*<sup>25</sup>; *PsCRISPA* (J. Hofer, personal communication).



**Figure 4** Complementation of *Arabidopsis as1-1* mutants by *SkARP1*. **a–h**, Vegetative phenotype of wild-type (**a**), *as1-1* (**b**) and 35S::*SkARP1*;*as1-1* (**c–h**) plants. **i**, Northern blot analysis of *SkARP1* transcript levels in *as1-1* (lane 1), *S. kraussiana* (lane 2) and 35S::*SkARP1*; *as1-1* (lanes 3–8) plants. Lanes 3–8 correspond to the lines shown in **c–h**, respectively. RNA was hybridized to the entire coding region of *SkARP1*. Arrows point to the endogenous *SkARP1* transcript in *S. kraussiana* and to the transgene transcript in *Arabidopsis*. Ethidium bromide fluorescence of ribosomal RNA is shown as a loading control.

proteins accumulate in *O. regalis* leaf primordia but not in mature leaves (Fig. 2d). We therefore propose that an overlap of *KNOX* and *ARP* expression in the meristem might facilitate shoot bifurcation whereas an overlap within leaf primordia confers meristematic properties on the leaf and enables the generation of complex leaf morphologies<sup>24</sup>.

To gain an overview of *KNOX*–*ARP* interactions in leaves, we mirrored *ARP* and class I *KNOX* phylogenies, and mapped *KNOX* expression patterns and the effect of loss of *ARP* function on to them (Fig. 3). Examining *KNOX* expression patterns in this context revealed parallel evolution of distinct patterns in the eudicot, monocot and lycophyte lineages. Eudicot *KNOX* expression patterns in one of the three main clades varied (clade 3), with some genes expressed in the central zone of the SAM and others in the rib zone. In a sister group (clade 2), two subclades were characterized by *KNOX* expression in either the central zone (CZ type) or the peripheral zone (PZ type). These two patterns predominate in monocots (clade 1), although their distribution varies between gene clades. Both patterns are also seen in *Selaginella* (Fig. 1). In the monocot maize (*Zea*), *ARP* function represses both CZ-type and PZ-type activities in the leaf. However, a different relationship emerges in eudicots. With the exception of *Antirrhinum*, loss of *ARP* function in eudicots leads to the ectopic expression of PZ-type but not CZ-type genes in the leaf. *ARP* function therefore constrains the activity of PZ-type but not CZ-type genes. Moreover, in *Arabidopsis* and tomato (*Lycopersicon*), CZ-type genes repress *ARP* expression, indicating a possible constraint in the opposite direction<sup>13,25</sup>. The *Antirrhinum* exception might reflect functional divergence or switching of roles between the two *ARP* duplicates.

The functional significance of CZ-type versus PZ-type patterns has been explained in several flowering plants<sup>20</sup>. In brief, CZ-type genes facilitate meristem formation and indeterminacy, whereas PZ-type genes facilitate the positioning of leaves and internode expansion. We propose that *ARP*-mediated repression of all *KNOX* genes in the leaf, as in maize, reflects the relatively simple and invariant leaf morphology found in the grasses. In contrast, the specialization of *KNOX*–*ARP* interactions in eudicots invokes a more complex repression mechanism, both between the meristem and the leaf and within leaf primordia. This complexity permits

multiple outputs that are reflected in the array of eudicot leaf diversity. We predict that *ARP* genes regulate both CZ-type and PZ-type activities in lycophytes, in which leaf morphology and anatomy are simple and invariant.

Despite subtle differences in the nature of *ARP*–*KNOX* interactions, the monocot *ARP* gene *rs2* complements the eudicot *arp* mutant *as1-1* (ref. 26). To determine the extent of functional equivalence between lycophyte and eudicot *ARP* genes, we introduced *SkARP1* into *as1-1* mutants under the control of a constitutive promoter. Compared with wild-type and *as1-1* controls (Fig. 4a, b), the phenotype of transformed plants varied from *as1*-like (Fig. 4c–e) to wild type (Fig. 4f–h). The degree of phenotypic rescue was inversely proportional to the number of transgenes (data not shown) and directly proportional to *SkARP1* transcript levels (Fig. 4i). These observations suggest that post-transcriptional silencing of the transgene occurred in lines that retained the *as1-1* mutant phenotype. The phenotype of non-silenced lines demonstrated that *SkARP1* is able to repress *KNOX* gene expression in the *Arabidopsis* leaf. In this context, *SkARP1* is therefore functionally equivalent to *AS1*.

The fossil record is insufficiently comprehensive for it to be stated unambiguously how many times leaves have evolved. However, it is generally accepted that determinate lateral outgrowths arose independently in ancestral lycophytes and euphyllophytes<sup>4</sup>. Within the euphyllophytes, it is not clear how ‘leaf-like’ these lateral outgrowths were. Consequently, it has been suggested that leaves evolved independently in the sphenopsids, ferns, progymnosperms and seed plants (that is, four times within the euphyllophytes). However, a morphometric analysis of more than 600 fossils implied that a ‘common developmental process underpins leaf development in all four groups<sup>3</sup>. *KNOX*–*ARP* interactions fulfil this role in ferns and seed plants, with spatial and temporal changes in gene expression leading to variations in leaf morphology and growth patterns. Our data indicate that the same developmental mechanism might operate in at least one lycophyte and therefore that this mechanism was recruited at least twice independently during land plant evolution. The evolution of leaves was thus constrained by a fundamental developmental programme that might originally have operated in primitive plants to regulate shoot branching. □

## Methods

### Plant materials

Plants were obtained from the Royal Botanic Gardens at Kew, the University of Oxford Botanic Gardens, local garden centres in Oxford, and the Nottingham Arabidopsis Stock Centre.

### Gene isolation

An *SkKNOX* fragment and *SkARP1* and *SvARP1* sequences were isolated by genomic polymerase chain reaction (PCR) with degenerate primers. Full-length sequences of *SkKNOX1*, *SkKNOX2* and *SkKNOX3* were obtained by screening a *Selaginella* shoot complementary DNA library with the PCR-amplified fragment.

### Phylogenetic analyses

To maximize sampling in *KNOX* and *MYB* phylogenies, conserved homeodomain and *MYB* sequences from *Arabidopsis KNOX* and *ARP* genes were used in tBLASTX searches of the following genome databases: <http://blast.genome.jp/>, [www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/), [www.plantdb.org/](http://www.plantdb.org/) and [www.moss.leeds.ac.uk/](http://www.moss.leeds.ac.uk/). Only EST or cDNA sequences were retrieved to avoid alignment problems with introns.

### Transgenic Arabidopsis

The full *SkARP1* coding sequence was amplified by PCR. The PCR product was inserted into the pART27 binary vector by using the pART7 shuttle vector. Thus, *SkARP1* was flanked by the cauliflower mosaic virus 35S promoter and a transcriptional terminator. Constructs were introduced into *Agrobacterium* strain GV3101 by electroporation and then into *Arabidopsis as1-1* mutant plants by floral dipping.

Received 7 December 2004; accepted 28 January 2005; doi:10.1038/nature03410.

- Edwards, D., Feehan, J. & Smith, D. G. A late Wenlock flora from County Tipperary, Ireland. *Bot. J. Linn. Soc.* **86**, 19–36 (1983).
- Gifford, E. M. & Foster, A. S. *Morphology and Evolution of Vascular Plants* (Freeman, New York, 1989).



3. Boyce, C. K. & Knoll, A. H. Evolution of developmental potential and the multiple independent origin of leaves in Paleozoic vascular plants. *Paleobiology* **28**, 70–100 (2002).
4. Kenrick, P. & Crane, P. R. *The Origin and Early Diversification of Land Plants: A Cladistic Study* (Smithsonian Institution Press, London, 1997).
5. Jackson, D., Veit, B. & Hake, S. Expression of maize *Knotted1* related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot. *Development* **120**, 404–413 (1994).
6. Lincoln, C., Long, J., Yamaguchi, J., Serikawa, K. & Hake, S. A. *Knotted1*-like homeobox gene in *Arabidopsis* is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. *Plant Cell* **6**, 1859–1876 (1994).
7. Nishimura, A., Tamaoki, M., Sato, Y. & Matsuoka, M. The expression of tobacco *Knotted1*-type class I homeobox genes correspond to regions predicted by the cytohistological zonation model. *Plant J.* **18**, 337–347 (1999).
8. Long, J. A., Moan, E. I., Medford, J. I. & Barton, M. K. A member of the KNOTTED class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. *Nature* **379**, 66–69 (1996).
9. Sentoku, N. *et al.* Regional expression of the rice *KN1*-type homeobox gene family during embryo, shoot and flower development. *Plant Cell* **11**, 1651–1663 (1999).
10. Hofer, J. M., Gourelay, C. W., Michael, A. & Ellis, T. H. Expression of a class I *Knotted1*-like homeobox gene is down-regulated in pea compound leaf primordia. *Plant Mol. Biol.* **45**, 387–398 (2001).
11. Hareven, D., Gutfinger, T., Parnis, A., Eshed, Y. & Lifschitz, E. The making of a compound leaf: genetic manipulation of leaf architecture in tomato. *Cell* **84**, 735–744 (1996).
12. Muller, J. *et al.* *In vitro* interactions between barley TALE homeodomain proteins suggest a role for protein-protein associations in the regulation of *Knox* gene function. *Plant J.* **27**, 13–23 (2001).
13. Byrne, M. E. *et al.* *ASYMMETRIC LEAVES1* mediates leaf patterning and stem cell function in *Arabidopsis*. *Nature* **408**, 967–971 (2000).
14. Schneeberger, R., Tsiantis, M., Freeling, M. & Langdale, J. A. The *rough sheath2* gene negatively regulates homeobox gene expression during maize leaf development. *Development* **125**, 2857–2865 (1998).
15. Tsiantis, M., Schneeberger, R., Golz, J. F., Freeling, M. & Langdale, J. A. The maize *rough sheath2* gene and leaf development programs in monocot and dicot plants. *Science* **284**, 154–156 (1999).
16. Timmermans, M. C. P., Hudson, A., Becraft, P. W. & Nelson, T. *ROUGH SHEATH2*: A myb protein that represses *Knox* homeobox genes in maize lateral organ primordia. *Science* **284**, 151–153 (1999).
17. Smith, L. G., Greene, B., Veit, B. & Hake, S. A dominant mutation in the maize homeobox gene, *Knotted-1*, causes its ectopic expression in leaf cells with altered fates. *Development* **116**, 21–30 (1992).
18. McHale, N. A. & Koning, R. E. *PHANTASTICA* regulates development of the adaxial mesophyll in *Nicotiana* leaves. *Plant Cell* **16**, 1251–1262 (2004).
19. Wikstrom, N., Savolainen, V. & Chase, M. W. Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. Lond. B* **268**, 2211–2220 (2001).
20. Tsiantis, M. & Hay, A. Comparative plant development: the time of the leaf? *Nature Rev. Genet.* **4**, 169–180 (2003).
21. Steeves, T. A. & Sussex, I. M. Studies on the development of excised leaves in sterile culture. *Am. J. Bot.* **44**, 665–673 (1957).
22. Bierhorst, D. W. On the stem apex, leaf initiation and early leaf ontogeny in filiclean ferns. *Am. J. Bot.* **64**, 125–152 (1977).
23. Bharathan, G. *et al.* Homologies in leaf form inferred from *KNOX1* gene expression during development. *Science* **296**, 1858–1860 (2002).
24. Kim, M., McCormick, S., Timmermans, M. & Sinha, N. The expression domain of *PHANTASTICA* determines leaflet placement in compound leaves. *Nature* **424**, 438–443 (2003).
25. Kim, M. *et al.* Reduced leaf complexity in tomato wiry mutants suggests a role for *PHAN* and *KNOX* genes in generating compound leaves. *Development* **130**, 4405–4415 (2003).
26. Theodoris, G., Inada, N. & Freeling, M. Conservation and molecular dissection of *ROUGH SHEATH2* and *ASYMMETRIC LEAVES1* function in leaf development. *Proc. Natl Acad. Sci. USA* **100**, 6837–6842 (2003).
27. Pautot, V. *et al.* *KNAT2*: evidence for a link between *Knotted*-like genes and carpel development. *Plant Cell* **13**, 1719–1734 (2001).
28. Golz, J. F., Keck, E. J. & Hudson, A. Spontaneous mutations in *KNOX* genes give rise to a novel floral structure in antirrhinum. *Curr. Biol.* **12**, 515–522 (2002).
29. Parnis, A. *et al.* The dominant developmental mutants of tomato, *Mouse-Ear* and *Curl*, are associated with distinct modes of abnormal transcriptional regulation of a *Knotted* gene. *Plant Cell* **9**, 2143–2158 (1997).
30. Janssen, B.-J., Lund, L. & Sinha, N. Overexpression of a homeobox gene *LeT6* reveals indeterminate features in the tomato compound leaf. *Plant Physiol.* **117**, 771–786 (1998).

Supplementary Information accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. Stork for technical help; C. Hammond and T. Gardner for help during their undergraduate degrees (funded by the Wellcome Trust and a Nuffield Bursary, respectively); M. Tsiantis and G. Theodoris for *rs2* fusion constructs; M. Tsiantis and M. Knight for criticizing our manuscript; M. Tsiantis, D. Barker and the Oxford Systematics Discussion Group for discussions; and J. Hofer, A. Hudson and N. McHale for the use of unpublished data presented in Fig. 3. D.L.A. was the recipient of a Sainsbury PhD studentship. The work was funded by a grant to J.A.L. and R.W.S. from the Biotechnology and Biological Sciences Research Council.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to J.A.L. ([jane.langdale@plants.ox.ac.uk](mailto:jane.langdale@plants.ox.ac.uk)). Sequences have been deposited in GenBank as follows: *SkKNOX1* (AY667449), *SkKNOX2* (AY667450), *SkKNOX3* (AY667451), *SkARPI* (AY667452) and *SvARPI* (AY667453).

## Leptin regulation of bone resorption by the sympathetic nervous system and CART

Florent Elefteriou<sup>1,2,3</sup>, Jong Deok Ahn<sup>1,2</sup>, Shu Takeda<sup>1,2,4,5</sup>, Michael Starbuck<sup>1,2</sup>, Xiangli Yang<sup>1,2</sup>, Xiyun Liu<sup>1,2</sup>, Hisataka Kondo<sup>5,6</sup>, William G. Richards<sup>7</sup>, Tony W. Bannion<sup>7</sup>, Masaki Noda<sup>5,6</sup>, Karine Clement<sup>8</sup>, Christian Vaisse<sup>9</sup> & Gerard Karsenty<sup>1,2,3</sup>

<sup>1</sup>Department of Molecular and Human Genetics, <sup>2</sup>Bone Disease Program of Texas, <sup>3</sup>Children's Nutrition Research Center, Baylor College of Medicine, Houston, Texas 77030, USA

<sup>4</sup>Department of Orthopedics, <sup>5</sup>Center of Excellence Program for Frontier Research on Molecular Destruction and Reconstruction of Tooth and Bone, <sup>6</sup>Department of Molecular Pharmacology, Medical Research Institute, Tokyo Medical and Dental University, Tokyo 101-0062, Japan

<sup>7</sup>Amgen Inc., Neuroscience, Thousand Oaks, California 91320, USA

<sup>8</sup>INSERM Avenir team—University Paris 6, EA3502, and CHRU Pitié Salpêtrière, Hôtel-Dieu Nutrition Department, F-75004 Paris, France

<sup>9</sup>Diabetes Center and Department of Medicine, University of California San Francisco, San Francisco, California 94143, USA

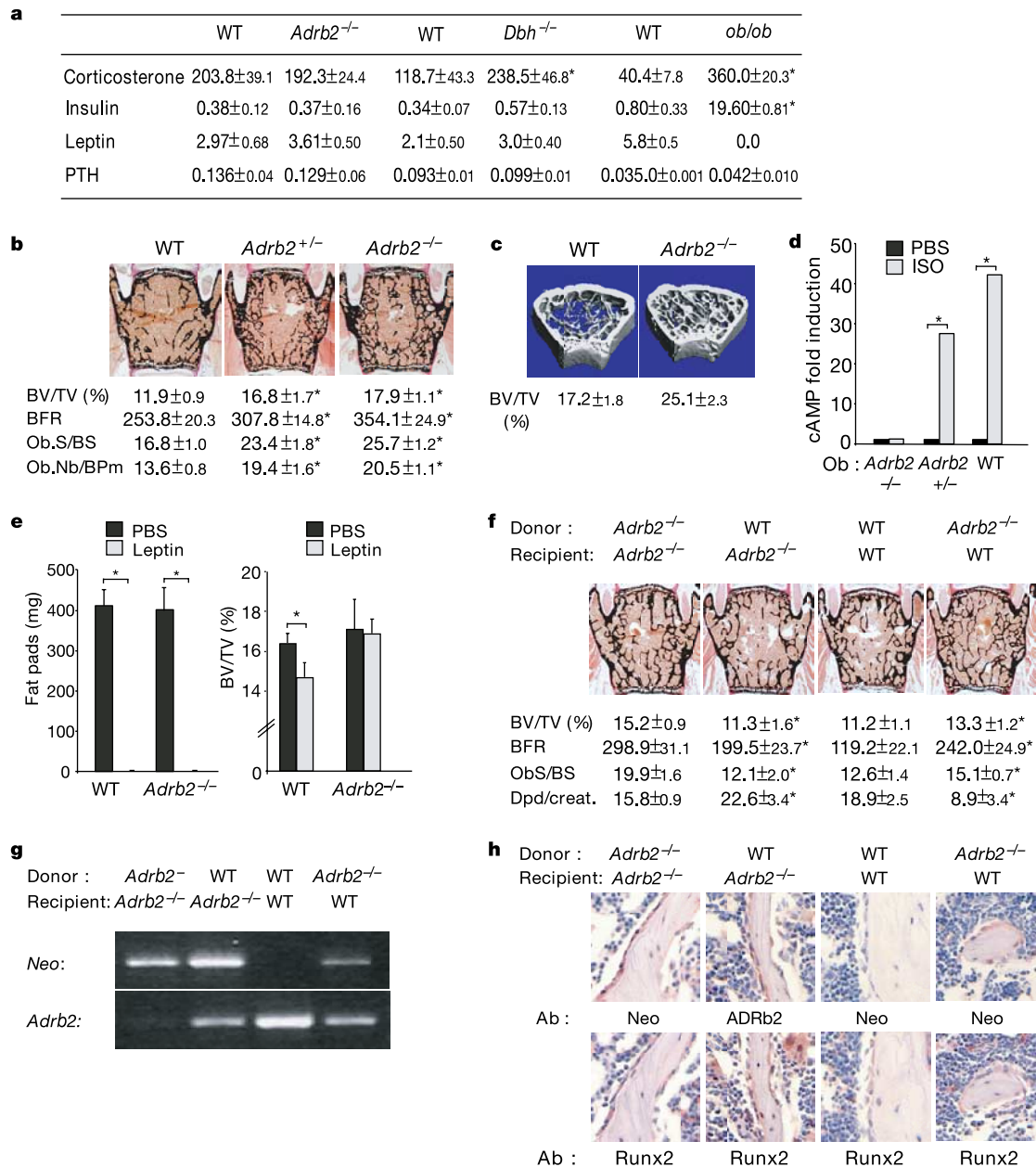
Bone remodelling, the mechanism by which vertebrates regulate bone mass, comprises two phases, namely resorption by osteoclasts and formation by osteoblasts; osteoblasts are multifunctional cells also controlling osteoclast differentiation. Sympathetic signalling via  $\beta$ 2-adrenergic receptors (*Adrb2*) present on osteoblasts controls bone formation downstream of leptin<sup>1</sup>. Here we show, by analysing *Adrb2*-deficient mice, that the sympathetic nervous system favours bone resorption by increasing expression in osteoblast progenitor cells of the osteoclast differentiation factor *Rankl*. This sympathetic function requires phosphorylation (by protein kinase A) of ATF4, a cell-specific CREB-related transcription factor essential for osteoblast differentiation and function<sup>2</sup>. That bone resorption cannot increase in gonadectomized *Adrb2*-deficient mice highlights the biological importance of this regulation, but also contrasts sharply with the increase in bone resorption characterizing another hypogonadic mouse with low sympathetic tone, the *ob/ob* mouse<sup>3</sup>. This discrepancy is explained, in part, by the fact that CART ('cocaine amphetamine regulated transcript'), a neuropeptide whose expression is controlled by leptin and nearly abolished in *ob/ob* mice<sup>4</sup>, inhibits bone resorption by modulating *Rankl* expression. Our study establishes that leptin-regulated neural pathways control both aspects of bone remodelling, and demonstrates that integrity of sympathetic signalling is necessary for the increase in bone resorption caused by gonadal failure.

Leptin antiosteogenic function is mediated by the sympathetic nervous system (SNS) acting through *Adrb2*, the only adrenergic receptor expressed in osteoblasts<sup>1</sup> (Supplementary Fig. 1). If both arms of bone remodelling are regulated by similar mechanisms, these results imply that bone resorption (BR) is controlled by neural means. To test this hypothesis, we used mutant mice in which pathways acting downstream of leptin signalling were disrupted.

*Adrb2*<sup>−/−</sup> mice have normal body weight and fat pad weight<sup>5</sup>, and none of the endocrine abnormalities observed in mice lacking leptin (*ob/ob*) or noradrenaline (*Dopamine-β-hydroxylase* (*Dbh*)<sup>−/−</sup> mice, Fig. 1a)<sup>6,7</sup>. Analyses of vertebrae and long bones revealed two unanticipated features in 6-month-old male and female *Adrb2*<sup>−/−</sup> mice. First, *Adrb2*<sup>−/−</sup> mice had a more severe high bone mass phenotype (HBM) than *ob/ob* or wild-type (WT) mice receiving  $\beta$ -blockers<sup>1,3</sup> (Fig. 1b, c). Illustrating the importance of sympathetic signalling in bone remodelling, this HBM also affected *Adrb2*<sup>+/−</sup>

mice yet *Adrb2*<sup>+/-</sup> osteoblasts do transduce a signal through *Adrb2* following treatment with isoproterenol (ISO), a surrogate of sympathetic signalling (Fig. 1b, d). This HBM was not observed in mice lacking *Adrb1* (Supplementary Fig. 2), indicating that *Adrb2*<sup>-/-</sup> mice are the best model to elucidate how sympathetic signalling in bone cells regulates bone mass. That long-term leptin intracerebroventricular (ICV) infusion did not reduce bone mass of *Adrb2*<sup>-/-</sup> mice established that SNS integrity is necessary for leptin antiosteogenic function (Fig. 1e).

To determine whether *Adrb2*<sup>-/-</sup> mice HBM involves bone cell-autonomous mechanisms, we performed transplantation of non-adherent bone marrow cells (BMCs)<sup>8</sup>. Transplantation of WT BMC into  $\gamma$ -irradiated *Adrb2*<sup>-/-</sup> mice normalized bone formation parameters; conversely, transplantation of *Adrb2*<sup>-/-</sup> BMC into  $\gamma$ -irradiated WT mice significantly increased bone formation (Fig. 1f). Polymerase chain reaction (PCR) analysis of *in vitro* differentiated osteoblastic colonies showed the presence of WT or *Adrb2*<sup>-/-</sup> osteoblasts in *Adrb2*<sup>-/-</sup> or WT  $\gamma$ -irradiated mice,



**Figure 1** Increased bone formation in *Adrb2*-deficient mice. **a**, Hormonal measurements (ng ml<sup>-1</sup>, mean + s.e.m.) in 6- (*Adrb2*<sup>-/-</sup> and *Dbh*<sup>-/-</sup>) and 3-month-old (*ob/ob*) mice. Differences in ages and genetic backgrounds explain different hormonal levels. **b**, Bone volume/tissue volume (BV/TV, %), bone formation rate (BFR,  $\mu\text{m}^3 \mu\text{m}^{-2} \text{yr}^{-1}$ ), osteoblast surface/bone surface (Ob.S/BS) and osteoblast number/bone perimeter (Ob.Nb/BPm) in 6-month-old mice (mean + s.e.m.). **c**,  $\mu\text{CT}$  analysis of 6-month-old distal femurs. **d**, cAMP production in phosphate buffered saline (PBS) and isoproterenol (ISO)-treated

osteoblasts (Ob). **e**, Fat pad weight and BV/TV following leptin ICV infusion. **f**, Formation and resorption parameters following transplantations. Comparison is between the same recipient genotype (mean + s.e.m.). **g**, PCR analysis of transplantation efficiency using Neomycin (Neo) and *Adrb2*-specific primers. **h**, Immunocytochemical detection of Neomycin<sup>+</sup>, Runx2<sup>+</sup> and *Adrb2*<sup>+</sup> osteoblasts in WT and *Adrb2*<sup>-/-</sup> bones following transplantation. Error bars, mean + s.e.m. \*Statistically significant.

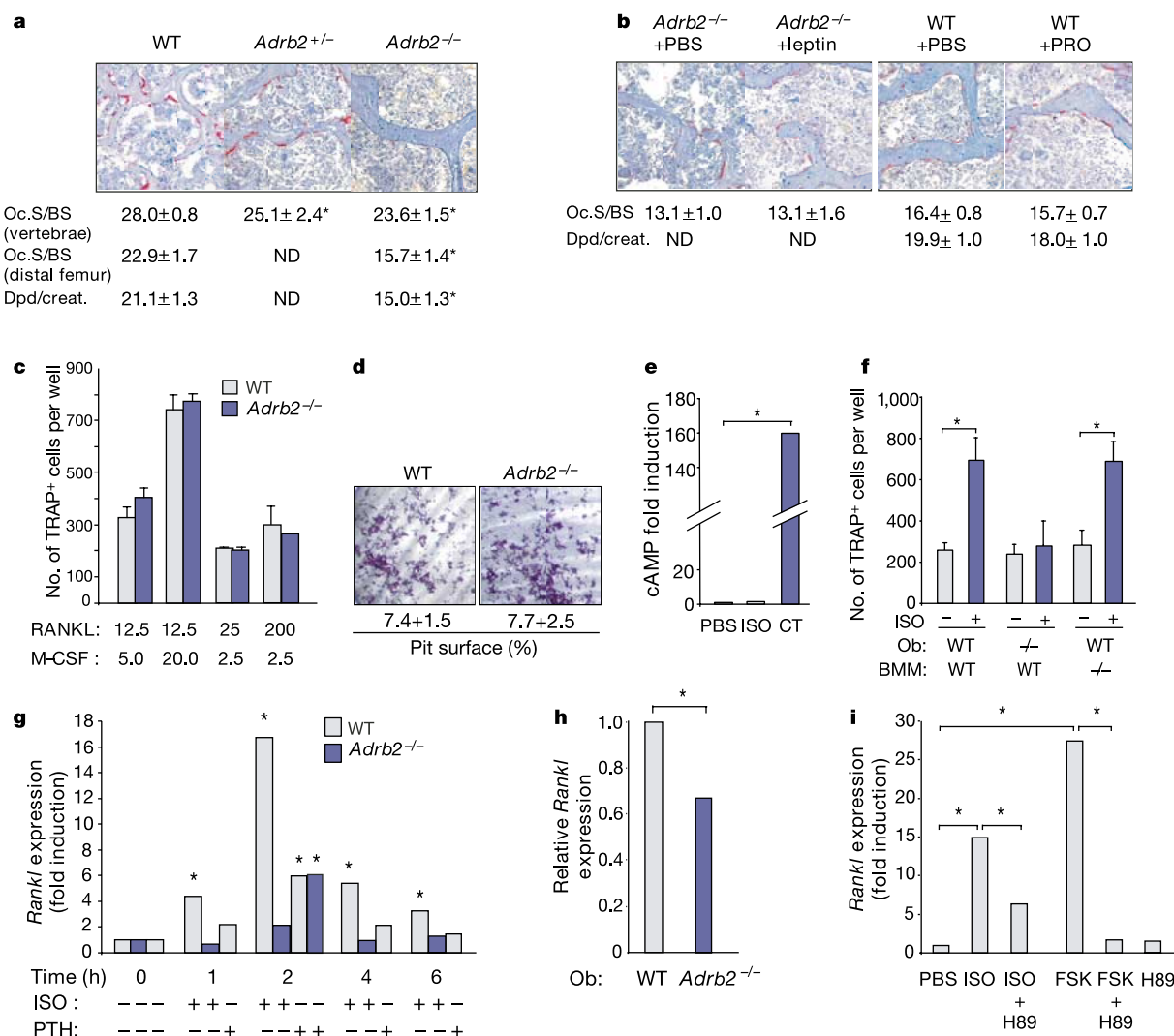
respectively (Fig. 1g and data not shown). Immunocytochemistry analysis showed Neomycin<sup>+</sup>, Runx2<sup>+</sup> osteoblasts in WT mice transplanted with *Adrb2*<sup>-/-</sup> BMCs and *Adrb2*<sup>+</sup>, Runx2<sup>+</sup>, Neomycin<sup>-</sup> osteoblasts in *Adrb2*<sup>-/-</sup> mice transplanted with WT BMCs (Fig. 1h). Thus the SNS controls bone mass by acting, at least in part, on cells of the osteoblast lineage.

Further cellular analysis of *Adrb2*<sup>-/-</sup> mice revealed a second unanticipated feature. Besides the expected increase in bone formation parameters<sup>1</sup> (Fig. 1b), there was, in all bones analysed, a significant decrease in BR parameters, including a decrease in number of tartrate-resistant acid phosphatase (TRAP)-positive multi-nucleated osteoclasts, indicative of a defect in osteoclast differentiation, and a decrease in urinary elimination of deoxy-pyridinoline (Dpd), a marker of osteoclast function (Fig. 2a). This was unexpected, because *ob/ob* or  $\beta$ -blocker-treated WT mice<sup>1</sup> have high and normal BR, respectively (Fig. 2b and data not shown). A decrease in Dpd urinary elimination was also observed in WT mice transplanted with *Adrb2*<sup>-/-</sup> BMC, while WT BMC transplantation into *Adrb2*<sup>-/-</sup> mice increased it (Fig. 1f). These BR abnormalities

were not corrected by leptin ICV infusion, indicating that leptin signalling regulates BR via the SNS (Fig. 2b).

To determine whether sympathetic signalling acts on osteoclast, we cultured bone marrow macrophages (BMMs) in the presence of RANKL, an osteoclast differentiation factor, and M-CSF, an osteoclast proliferation factor<sup>9,10</sup>. WT or *Adrb2*<sup>-/-</sup> BMMs differentiated equally well into osteoclasts at each RANKL/M-CSF dose tested (Fig. 2c), and ISO did not hamper generation of osteoclasts (Supplementary Fig. 3). *Adrb2* inactivation did not alter the ability of osteoclasts to generate resorption pits on dentine slices, and ISO did not increase cAMP production in osteoclasts (Fig. 2d-e).

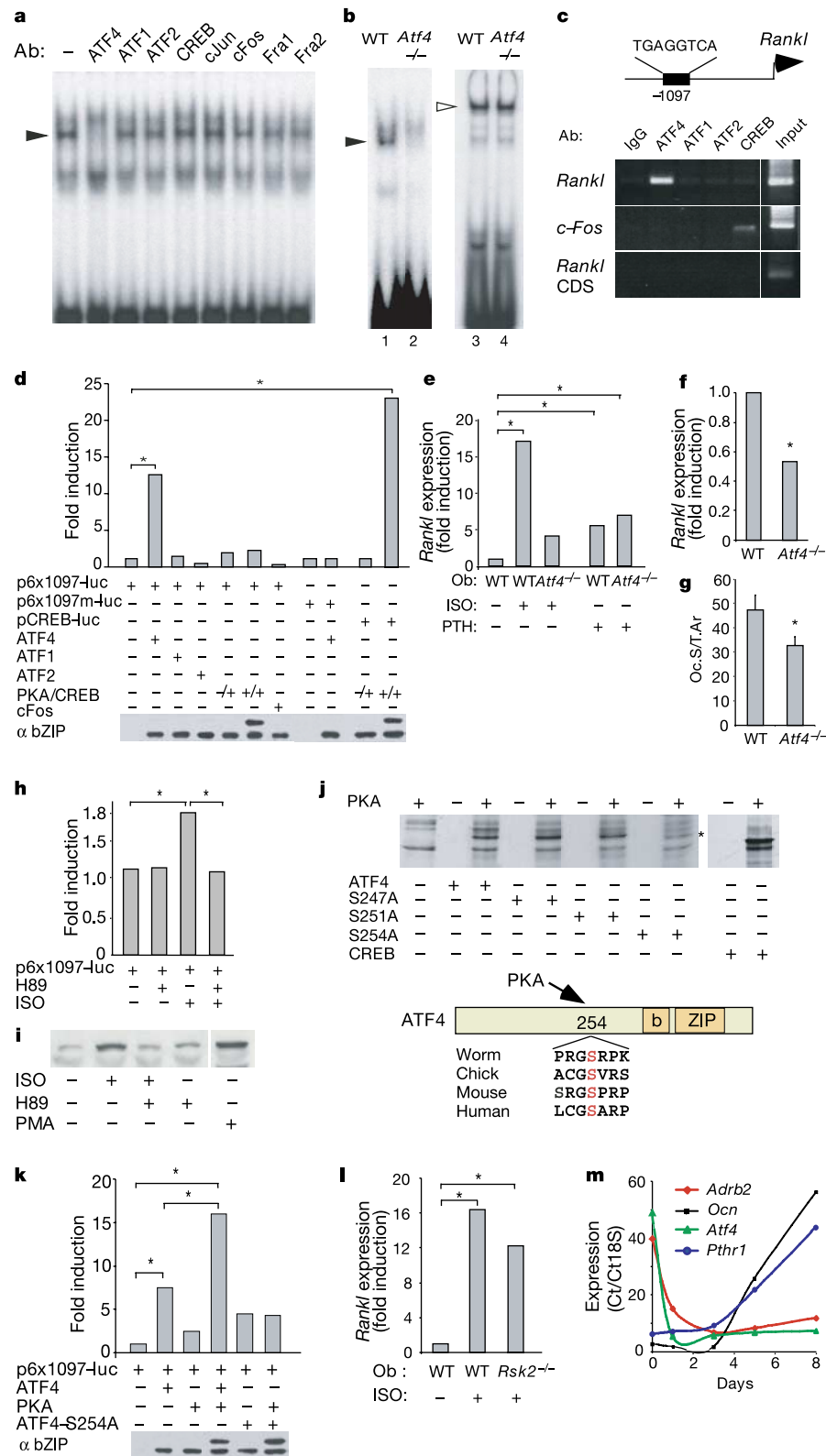
Next we asked whether the SNS affects BR by acting in osteoblasts, a cell type controlling osteoclast differentiation<sup>9</sup>. ISO enhanced generation of osteoclasts when WT, but not *Adrb2*<sup>-/-</sup>, osteoblasts were co-cultured with WT BMMs (Fig. 2f). Moreover, ISO treatment of WT, but not *Adrb2*<sup>-/-</sup>, osteoblasts increased expression of *Rankl* (a secreted osteoclast differentiation factor) to a greater extent than did parathyroid hormone (PTH), a hormone



**Figure 2** Sympathetic signalling in osteoblasts regulates BR. **a,b**, Osteoclast surface/bone surface and Dpd/creatinine in 6-month-old mice; **b**, following leptin ICV or propranolol (PRO) treatment. ND, not determined. **c**, *In vitro* differentiation of WT and *Adrb2*<sup>-/-</sup> BMMs with limiting amounts of RANKL and M-CSF (ng ml<sup>-1</sup>). **d**, WT and *Adrb2*<sup>-/-</sup> osteoclasts generate resorption pits equally well. **e**, Calcitonin, not ISO,

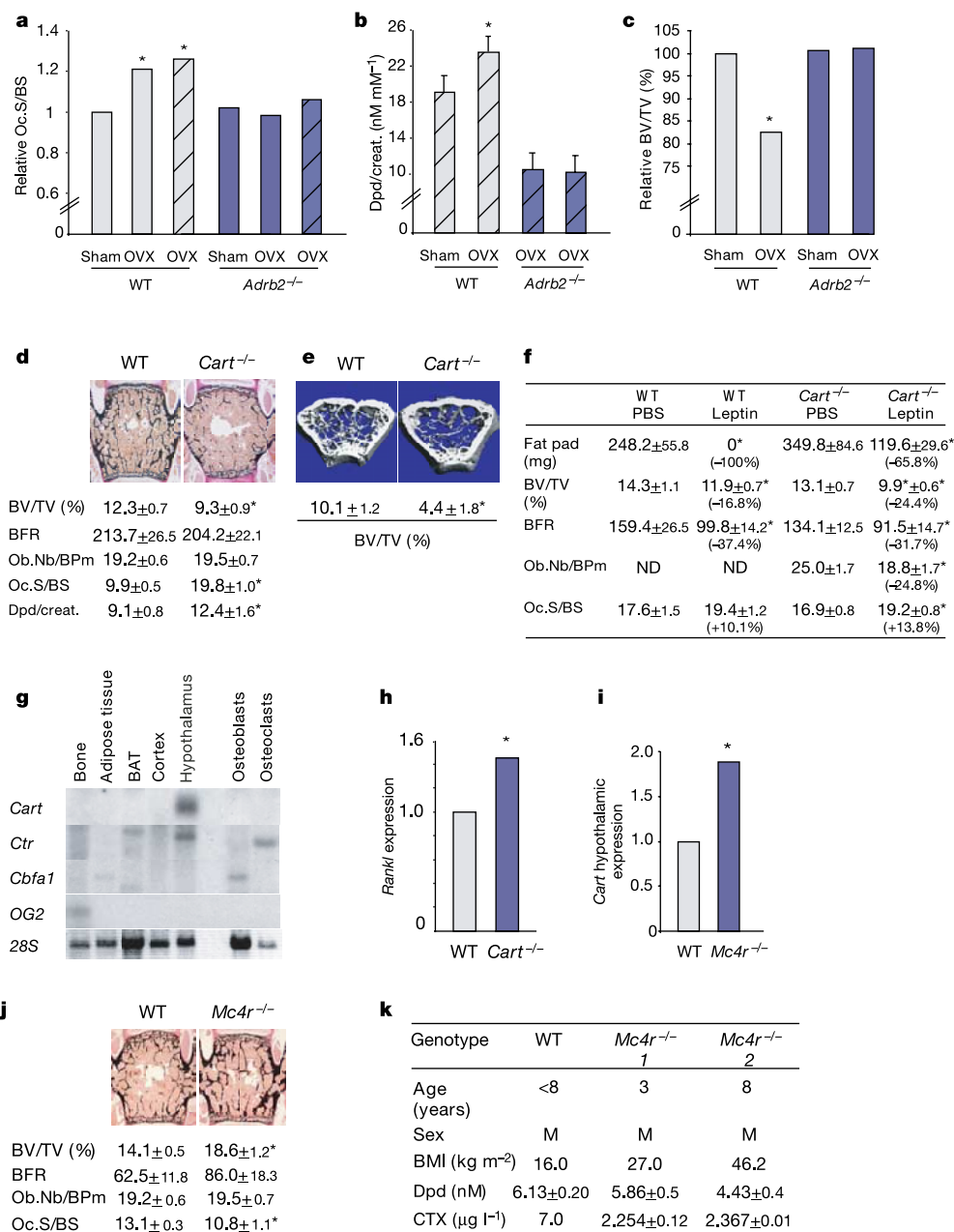
induces cAMP production in osteoclasts. **f**, Number of TRAP<sup>+</sup> cells following co-culture of osteoblasts and BMMs in the presence of ISO. **g-i**, Real-time PCR. ISO induces *Rankl* expression in WT not *Adrb2*<sup>-/-</sup> osteoblasts. *Rankl* expression is decreased in *Adrb2*<sup>-/-</sup> osteoblasts and PKA is required for ISO induction of *Rankl* expression (*n* = 3 per experiment). Error bars, mean + s.e.m. \*Statistically significant.





**Figure 3** Sympathetic regulation of *Rankl* expression. **a**, EMSA: ATF4 antibody prevented formation of protein–DNA complex (filled arrow) on *Rankl* promoter CRE site. **b**, WT (1) but not *Atf4*<sup>-/-</sup> (2) osteoblast nuclear extracts bind to -1097<sup>Rankl</sup>. Sp1 (open arrow) served as control (3, 4). **c**, ChIP: ATF4 binds to -1097<sup>Rankl</sup> but not to *Rankl* coding sequence (CDS). **d**, DNA cotransfections in COS cells using multimers of WT or mutant -1097<sup>Rankl</sup> (p6x1097-luc) or pCREB-luc and indicated expression vectors. **e**, Real-time PCR. *Rankl* induction by ISO but not by PTH is blunted in *Atf4*<sup>-/-</sup> osteoblasts. **f, g**, Decreased *Rankl* expression and Oc.S/Tar in absence of *Atf4*<sup>-/-</sup>. **h**, ISO increased

ATF4 transactivation activity, H89 inhibited it, in ROS 17/2.8 cells. **i**, Phospho-<sup>254</sup>ATF4 Western blot. H89 inhibits ATF4 phosphorylation following ISO treatment. **j**, *In vitro* kinase assay. ATF4 is phosphorylated by PKA on serine 254 (asterisk). CREB served as positive control. **k**, DNA cotransfection assays in COS cells. PKA increases ATF4 transactivation activity through serine 254. **l, m**, Real-time PCR. ISO induces *Rankl* expression in *Rsk2*<sup>-/-</sup> osteoblasts. *Adrb2*, *Pthr*, *Atf4* and *Osteocalcin* (*Ocn*) expression during osteoblast differentiation. Bottom line of **d** and **k** shows Western blots of transfected proteins. \*Statistically significant.



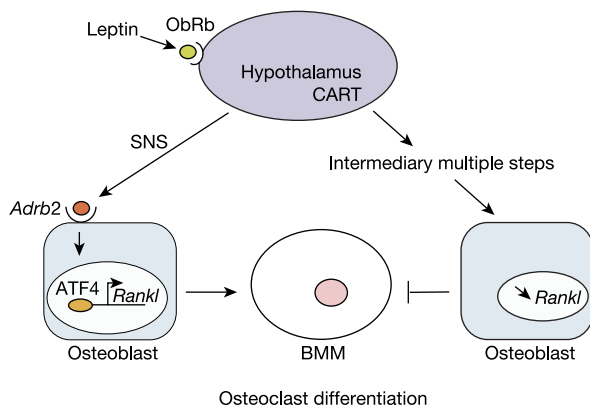
**Figure 4** SNS and CART antagonistic functions. **a–c**, Relative Oc.S/BS, urinary Dpd and BV/TV in WT and *Adrb2*<sup>-/-</sup> mice ovariectomized (OVX) for 4 or 12 weeks (plain or hatched bars). \**P* < 0.05 versus WT sham littermates. **d**, BV/TV, formation and resorption parameters in 6-month-old *Cart*<sup>-/-</sup> mice. **e**,  $\mu$ CT analysis of 6-month-old femurs. **f**, Leptin ICV infusion in WT and *Cart*<sup>-/-</sup> mice. *Cart* deletion enhances leptin antiosteogenic and proresorptive functions. **g**, Northern blots. *Cart* is not expressed in

bone cells. **h,i**, Real-time PCR. Increased *Rankl* and *Cart* expression in *Cart*<sup>-/-</sup> and *Mc4r*<sup>-/-</sup> bones and hypothalami. **j**, Increased BV/TV, normal bone formation, decreased bone resorption in 6-month-old *Mc4r*<sup>-/-</sup> mice. **k**, Serum Dpd and crosslaps CTX are decreased in MC4R-deficient patients. Error bars, mean + s.e.m. BMI, body mass index.

that upregulates *Rankl* expression in both WT<sup>9</sup> and *Adrb2*<sup>-/-</sup> osteoblasts (Fig. 2g). Accordingly, *Rankl* expression was decreased in *Adrb2*<sup>-/-</sup> osteoblasts (Fig. 2h). ISO did not affect expression of *Osteoprotegerin* (a *Rankl* decoy receptor) or of other cytokines (Supplementary Fig. 4).

ISO induction of *Rankl* was blunted by a PKA inhibitor, suggesting that CREB mediates it (Fig. 2i). A CREB-responsive element is present in *Rankl* promoter at -1097, and a protein–DNA complex formed upon incubation of osteoblast nuclear extracts

with this element in electrophoretic mobility shift assay (EMSA) (Fig. 3a). Surprisingly, this protein–DNA complex was not affected by an antibody against CREB but was abolished by an antibody against ATF4, an osteoblast-specific CREB/ATF family member essential for osteoblast function<sup>2</sup>. Several lines of evidence demonstrated that ATF4 mediates sympathetic regulation of *Rankl* expression in osteoblasts. First, in EMSA, this protein–DNA complex did not form when nuclear extracts of *Atf4*<sup>-/-</sup> osteoblasts were used as a source of proteins (Fig. 3b); second, in chromatin



**Figure 5** Model of the neuronal control of bone resorption.

immunoprecipitation, ATF4 bound to this sequence while CREB did not, although it bound to the *c-Fos* promoter, a known target gene<sup>11</sup> (Fig. 3c). Third, in DNA cotransfection experiments, ATF4 transactivated a vector containing 6 copies of the WT but not of a mutant  $-1097^{\text{Rankl}}$  promoter element; CREB and other leucine zipper proteins did not (Fig. 3d); fourth, unlike PTH, isoproterenol did not upregulate *Rankl* expression in *Atf4*<sup>-/-</sup> osteoblasts (Fig. 3e); fifth, *Rankl* expression and osteoclast surface were decreased in *Atf4*<sup>-/-</sup> osteoblasts and bones, respectively (Fig. 3f, g).

ISO increased ATF4 transactivation function and triggered its phosphorylation, two events prevented by PKA inhibition (Fig. 3h, i). A consensus PKA phosphorylation site exists in ATF4 at serine 254. Mutating this highly conserved serine to alanine prevented PKA phosphorylation of ATF4 (Fig. 3j), and PKA forced expression enhanced ATF4 transactivation function (Fig. 3k). Conversely, ISO induced *Rankl* expression in osteoblasts lacking RSK2, another kinase regulating ATF4 function<sup>2</sup> (Fig. 3l), and failed to phosphorylate RSK2 (Supplementary Fig. 5). Thus, ATF4 phosphorylation by PKA but not by RSK2 is required for sympathetic regulation of *Rankl* expression.

ATF4 is required for sympathetic, but not PTH, regulation of *Rankl* expression, suggesting that sympathetic and PTH signalling target different subsets of osteoblasts. Indeed, *Adrb2* was expressed at its highest level in undifferentiated osteoblasts (day 0 of culture), when *Atf4* expression peaks, while *PTH receptor* expression peaked later in differentiated osteoblasts (Fig. 3m).

The biological importance of the sympathetic regulation of BR was addressed by ovariectomy of 1-month-old mice and histological analyses 4 and 12 weeks later. Ovariectomy increased osteoclast surface and Dpd urinary elimination in WT but not in *Adrb2*<sup>-/-</sup> mice. As a result, bone mass did not decrease in *Adrb2*<sup>-/-</sup> mice following gonadectomy (Fig. 4a–c). Although these results highlighted the importance of SNS integrity for the development of gonadal failure-induced bone loss, they were totally unexpected. Indeed, gonadectomized *Adrb2*<sup>-/-</sup> mice should be, in term of bone biology, a phenocopy of *ob/ob* mice that are hypogonadic and have a low sympathetic activity. Yet while *ob/ob* have an increase in BR, gonadectomized *Adrb2*<sup>-/-</sup> mice do not. This discrepancy implied that expression of gene(s) regulating BR is perturbed in *ob/ob* but not in gonadectomized *Adrb2*<sup>-/-</sup> mice.

To identify such inhibitors, we focused on genes whose expression is regulated by leptin but whose inactivation does not affect appetite or fertility. *Cart* encodes a neuropeptide whose expression in brain is increased by leptin, directly or indirectly, low in *ob/ob* mice<sup>3</sup> and normal in *Adrb2*<sup>-/-</sup> mice (Supplementary Fig. 6). *Cart*<sup>-/-</sup> mice have no overt phenotypic abnormalities<sup>12</sup>

(Supplementary Fig. 7) but displayed, in both sexes, a low bone mass phenotype at 6 months of age (Fig. 4d–e and Supplementary Fig. 8). Osteoblast numbers and bone formation rates were normal while osteoclast surface and number was nearly doubled in *Cart*<sup>-/-</sup> bones (Fig. 4d). The significant increase in urinary Dpd elimination established that *Cart*<sup>-/-</sup> osteoclasts were functional.

To determine if leptin-dependent sympathetic regulation of bone mass occurred in the absence of CART, we performed leptin ICV infusion in 1-month-old *Cart*<sup>-/-</sup> mice, a procedure increasing sympathetic signalling<sup>13</sup>. This infusion decreased bone mass and osteoclast surface more efficiently in *Cart*<sup>-/-</sup> mice than in WT littermates (Fig. 4f), demonstrating that leptin-mediated sympathetic regulation of bone mass is not impaired in the absence of CART. The more severe decrease of bone mass observed in *Cart*<sup>-/-</sup> mice following leptin ICV further established that CART is an inhibitor of BR. *Cart* is not expressed in bone cells (Fig. 4g), WT and *Cart*<sup>-/-</sup> BMMs differentiated equally well into osteoclasts, exogenous CART did not affect BMM differentiation into osteoclasts, and co-culture experiments failed to detect any *Cart*<sup>-/-</sup> cell-autonomous defect (Supplementary Fig. 9)—yet *Rankl* expression was upregulated in *Cart*<sup>-/-</sup> bones, suggesting that CART exerts its function by ultimately modulating *Rankl* signalling (Fig. 4h).

*Mc4r*<sup>-/-</sup> mice displayed an increase in hypothalamic *Cart* expression and, at 6 months of age, a HBM (Fig. 4i, j). While bone formation parameters were normal in *Mc4r*<sup>-/-</sup> mice, excluding the possibility that their HBM was secondary to a dysfunction of the leptin-dependent sympathetic regulatory loop, they displayed a marked reduction in osteoclast number. This feature provided an explanation for their HBM and was consistent with the increase in *Cart* expression. This finding suggested that the increased bone density observed in *MC4R*-deficient patients<sup>14</sup> may be caused, at least in part, by a decrease in bone resorption. The decrease in serum levels of two BR markers, Dpd and CTX, in patients lacking *MC4R* verified this hypothesis (Fig. 4k).

Thus leptin controls BR through, at least, two distinct and antagonistic pathways. On one hand, sympathetic signalling via *Adrb2* promotes osteoclast differentiation; on the other hand, CART inhibits it. Although both pathways regulate *Rankl* expression, the molecular bases of CART regulation of BR remain elusive in absence of a specific CART receptor (Fig. 5). That sympathetic regulation of bone mass occurs in the absence of CART suggests that CART uses other means to regulate bone resorption. This hypothesis is further supported by two suggestive lines of evidence: first, CART and sympathetic signalling have opposite effects on BR; second, CART (unlike sympathetic signalling) has no detectable effect on bone formation. From a biomedical perspective, the increase in bone formation and the decrease in BR characterizing *Adrb2*<sup>-/-</sup> mice add significant credence to the contention that an efficient, bone-specific, pharmacological blockade of *Adrb2* signalling would be a great asset in the management of gonadal failure-induced bone loss. □

## Methods

### Animals, surgical procedures and histology

Mutant mice used in these studies have been described<sup>15,16</sup>. ICV infusions and propranolol treatment were performed as described<sup>1</sup> in 2- (*Adrb2*<sup>-/-</sup>) or 1- (*Cart*<sup>-/-</sup>) month-old mice. For bone marrow transplantations, 2-month-old mice lethally irradiated with 1,100 rad in a double dose were injected with  $2 \times 10^6$  nucleated whole bone marrow cells and killed 4 months later ( $n = 5-8$  per group). Bone marrow cells were then flushed from long bones, and differentiated *in vitro* in the presence of ascorbic acid ( $50 \mu\text{g ml}^{-1}$ ) for 10 days. DNA extracted from osteoblasts was used for PCR genotyping. Immunocytochemistry studies were performed on decalcified sections using commercially available Neomycin and *Adrb2* antibodies. Bone histology and histomorphometry analyses were performed as previously described<sup>13</sup>. Twelve- $\mu\text{m}$ -resolution micro-computed tomography ( $\mu\text{CT}$ ) measurements were performed on distal femurs. Five to ten mice were analysed for each group.



## Cell and molecular studies

Osteoblasts were used<sup>3</sup> at day 0 for ISO induction. Osteoclasts were differentiated with RANKL (50 ng ml<sup>-1</sup>) and M-CSF (30 ng ml<sup>-1</sup>). Co-cultures of osteoblasts/osteoclasts were performed as described<sup>17</sup>. Following treatment, multinucleated TRAP<sup>+</sup> cells were counted in triplicate wells. For pit resorption analysis, BMMs were cultured for 3 days with M-CSF and RANKL, trypsinized, plated on dentine slices, resorption pits stained with haematoxylin and the resorbed area quantified. Gene expression was assessed by real-time PCR on osteoblasts or tissue RNA. Nuclear extracts were prepared and EMSAs were performed as described<sup>18</sup>. Chromatin immunoprecipitation assays (ChIP) were performed using primary osteoblasts. Phospho-<sup>254</sup>ATF4 antibody was generated against NLPSPGGSRGSPPK peptide in which the underlined serine was phosphorylated. Cells were treated with ISO (10  $\mu$ M), PTH (1-34)(10 nM), forskolin (10  $\mu$ M), PMA (phorbol 12-myristate 13-acetate, 0.2  $\mu$ g ml<sup>-1</sup>) or H89 (30  $\mu$ M, 30 min pre-treatment). *In vitro* kinase assays were performed as described<sup>2</sup>. COS cells were transfected with 100 ng of reporter, 30–100 ng of expression plasmids and 15 ng of RSV- $\beta$ -gal reporter vector. ROS 17/2.8 were transfected as described<sup>18</sup>. Transfections were repeated at least 4 times in triplicates.

## Biochemistry

cAMP, Dpd crosslinks, creatinine urinary values, carboxy-terminal telopeptides of type-I collagen (CTX), leptin, insulin and PTH serum levels were measured using commercial kits.

## Statistical analyses

Data are expressed as mean  $\pm$  s.e.m. Statistical significance was assessed by Student's test. Values were considered statistically significant at  $P < 0.05$ .

Received 15 December 2004; accepted 25 January 2005; doi:10.1038/nature03398.  
Published online 20 February 2005.

1. Takeda, S. *et al.* Leptin regulates bone formation via the sympathetic nervous system. *Cell* **111**, 305–317 (2002).
2. Yang, X. *et al.* ATF4 is a substrate of RSK2 and an essential regulator of osteoblast biology; implication for Coffin-Lowry Syndrome. *Cell* **117**, 387–398 (2004).
3. Ducy, P. *et al.* Leptin inhibits bone formation through a hypothalamic relay: A central control of bone mass. *Cell* **100**, 197–207 (2000).
4. Kristensen, P. *et al.* Hypothalamic CART is a new anorectic peptide regulated by leptin. *Nature* **393**, 72–76 (1998).
5. Chruscinski, A. J. *et al.* Targeted disruption of the beta2 adrenergic receptor gene. *J. Biol. Chem.* **274**, 16694–16700 (1999).
6. Thomas, S. A., Matsumoto, A. M. & Palmiter, R. D. Noradrenaline is essential for mouse fetal development. *Nature* **374**, 643–646 (1995).
7. Friedman, J. M. & Halaas, J. L. Leptin and the regulation of body weight in mammals. *Nature* **395**, 763–770 (1998).
8. Dominici, M. *et al.* Hematopoietic cells and osteoblasts are derived from a common marrow progenitor after bone marrow transplantation. *Proc. Natl Acad. Sci. USA* **101**, 11761–11766 (2004).
9. Teitelbaum, S. L. & Ross, F. P. Genetic regulation of osteoclast development and function. *Nature Rev. Genet.* **4**, 638–649 (2003).
10. Lacey, D. L. *et al.* Osteoprotegerin ligand is a cytokine that regulates osteoclast differentiation and activation. *Cell* **93**, 165–176 (1998).
11. Berkowitz, L. A., Riabowol, K. T. & Gilman, M. Z. Multiple sequence elements of a single functional class are required for cyclic AMP responsiveness of the mouse c-fos promoter. *Mol. Cell. Biol.* **9**, 4272–4281 (1989).
12. Asnicar, M. A. *et al.* Absence of cocaine- and amphetamine-regulated transcript results in obesity in mice fed a high caloric diet. *Endocrinology* **142**, 4394–4400 (2001).
13. Satoh, N. *et al.* Sympathetic activation of leptin via the ventromedial hypothalamus: leptin-induced increase in catecholamine secretion. *Diabetes* **48**, 1787–1793 (1999).
14. Orwoll, B., Bouxsein, M. L., Marks, D. L., Cone, R. D. & Klein, R. F. in *ORS/AAOS Presentations 2003, 71st Annual Meeting of the AAOS (ORS, San Francisco, CA, 2004)*.
15. Rohrer, D. K., Chruscinski, A., Schauble, E. H., Bernstein, D. & Kobilka, B. K. Cardiovascular and metabolic alterations in mice lacking both beta1- and beta2-adrenergic receptors. *J. Biol. Chem.* **274**, 16701–16708 (1999).
16. Huszar, D. *et al.* Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* **88**, 131–141 (1997).
17. Takahashi, N. *et al.* Osteoblastic cells are involved in osteoclast formation. *Endocrinology* **123**, 2600–2602 (1988).
18. Ducy, P. & Karsenty, G. Two distinct osteoblast-specific cis-acting elements control expression of a mouse osteocalcin gene. *Mol. Cell. Biol.* **15**, 1858–1869 (1995).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank T. Townes and A. Hanauer for *Atf4* and *Rsk2*  $-/-$  mice, A. Hanauer and M. Montminy for RSK2 and CREB antibodies, M. Huelskamp and D.A. Horst for Dpd and CTX measurements, P. Ducy for suggestions and critical readings of the manuscript, and L. Li for technical assistance. This work was supported by grants from NIH, NSBRI and CNRC (G.K., E.E.), Arthritis Foundation (S.T.) and Children's Brittle Bone Foundation (X.Y.).

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to G.K. (karsenty@bcm.tmc.edu).

# Recognition of bacterial glycosphingolipids by natural killer T cells

Yuki Kinjo<sup>1</sup>, Douglass Wu<sup>2</sup>, Gisen Kim<sup>1</sup>, Guo-Wen Xing<sup>2</sup>, Michael A. Poles<sup>3,4</sup>, David D. Ho<sup>4</sup>, Moriya Tsuji<sup>4,5</sup>, Kazuyoshi Kawahara<sup>6,\*</sup>, Chi-Huey Wong<sup>2</sup> & Mitchell Kronenberg<sup>1</sup>

<sup>1</sup>Division of Developmental Immunology, La Jolla Institute for Allergy and Immunology, San Diego, 10355 Science Center Drive, San Diego, California 92121, USA

<sup>2</sup>Department of Chemistry and Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA

<sup>3</sup>Department of Medicine, Division of Gastroenterology, New York University School of Medicine, 550 First Avenue, New York, New York 10016, USA

<sup>4</sup>Aaron Diamond AIDS Research Center, The Rockefeller University, 455 First Avenue, New York, New York 10016, USA

<sup>5</sup>Department of Medical and Molecular Parasitology, New York University School of Medicine, 341 E. 25th Street, New York, New York 10010, USA

<sup>6</sup>The Kitasato Institute, Minato-ku, Tokyo 108-8642, Japan

\* Present address: Department of Applied Material and Life Science, College of Engineering, Kanto Gakuin University, Yokohama 236-8501, Japan

Natural killer T (NKT) cells constitute a highly conserved T lymphocyte subpopulation that has the potential to regulate many types of immune responses through the rapid secretion of cytokines<sup>1,2</sup>. NKT cells recognize glycolipids presented by CD1d, a class I-like antigen-presenting molecule. They have an invariant T-cell antigen receptor (TCR)  $\alpha$ -chain, but whether this invariant TCR recognizes microbial antigens is still controversial. Here we show that most mouse and human NKT cells recognize glycosphingolipids from *Sphingomonas*, Gram-negative bacteria that do not contain lipopolysaccharide<sup>3–5</sup>. NKT cells are activated *in vivo* after exposure to these bacterial antigens or bacteria, and mice that lack NKT cells have a marked defect in the clearance of *Sphingomonas* from the liver. These data suggest that NKT cells are T lymphocytes that provide an innate-type immune response to certain microorganisms through recognition by their antigen receptor, and that they might be useful in providing protection from bacteria that cannot be detected by pattern recognition receptors such as Toll-like receptor 4.

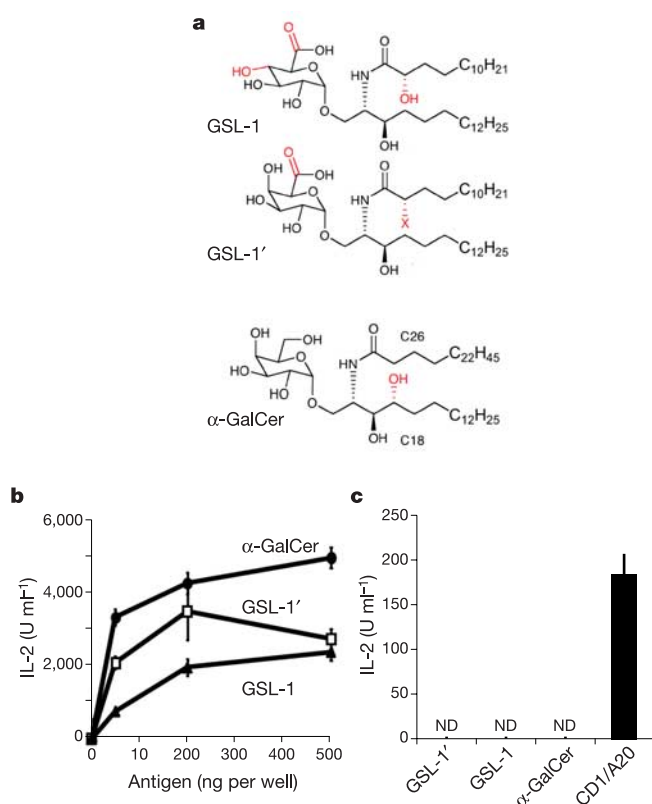
A glycosphingolipid, known as KRN 7000 or  $\alpha$ -galactosyl ceramide ( $\alpha$ -GalCer), was isolated from a marine sponge in a screen for compounds that could prevent tumour metastases to the liver of mice<sup>6</sup>. This activity was subsequently shown to be caused by the CD1d-mediated stimulation of NKT cells with an invariant V $\alpha$ 14 (V $\alpha$ 14i) TCR rearrangement<sup>7</sup>. Although in mice these form the majority of T cells that also express natural killer receptors, such as NK1.1, to distinguish them from other T cells that express natural killer receptors we refer to them as V $\alpha$ 14i NKT cells. The reactivity of NKT cells for  $\alpha$ -GalCer presented by CD1d is unusually conserved, because mouse V $\alpha$ 14i NKT cells recognize  $\alpha$ -GalCer presented by heterologous human CD1d<sup>8</sup>, and human NKT cells with an invariant rearrangement of V $\alpha$ 24 (V $\alpha$ 24i), the orthologue of mouse V $\alpha$ 14, recognize  $\alpha$ -GalCer presented by mouse CD1d<sup>8</sup>. Very few other examples of this type of interspecies cross-reactivity for T cells exist, indicating that this specificity and the conserved, invariant V $\alpha$  (V $\alpha$ i) rearrangements required for it might be particularly important. Although glycosphingolipids are widely distributed,  $\alpha$ -GalCer is unusual because it has an  $\alpha$  linkage of the 1' carbon of the sugar to the 1 carbon of the sphingosine base. However, in nearly all other glycosphingolipids the bond connecting the sugar to the lipid is in the  $\beta$  anomeric form, but this  $\beta$  form is not antigenic. It is unlikely that mice and humans have T-cell

populations selected for the recognition of marine sponge antigens. NKT cells exhibit a degree of autoreactivity, and recently it was shown that isoglobotrihexosyl ceramide (iGb3), a mammalian glycosphingolipid, can activate mouse V $\alpha$ 14i and human V $\alpha$ 24i NKT cells<sup>9</sup>, referred to collectively here as iNKT cells.

To determine whether microbial antigens that can activate most iNKT cells exist, we examined compounds from *Sphingomonas* bacteria, which are found in soil, sea water and plants, and are highly abundant in the environment<sup>10</sup>. Previously, we and others purified abundant glycosphingolipids from *Sphingomonas* species, including *S. paucimobilis* and *S. yanoikuyae*<sup>3–5</sup>. Two of these glycosphingolipids are shown in Fig. 1a; the differences between these compounds and  $\alpha$ -GalCer are highlighted. These compounds contain  $\alpha$ -linked glucuronic (GSL-1) or galacturonic (GSL-1') acid, and they are mixtures containing at least three different sphingosine bases, having either 18 (Fig. 1a), 20 or 21 carbons. Figure 1b indicates that the purified GSL-1 and GSL-1' compounds stimulated a V $\alpha$ 14i NKT cell hybridoma to release interleukin (IL)-2 when added to CD1d-coated plates, but they did not stimulate a CD1d-reactive hybridoma not expressing V $\alpha$ 14 and therefore not reactive to  $\alpha$ -GalCer (Fig. 1c), nor a peptide-reactive V $\beta$ 8.2<sup>+</sup> hybridoma (data not shown), showing the specificity of the GSL-1 stimulation. Neither compound was as active as synthetic

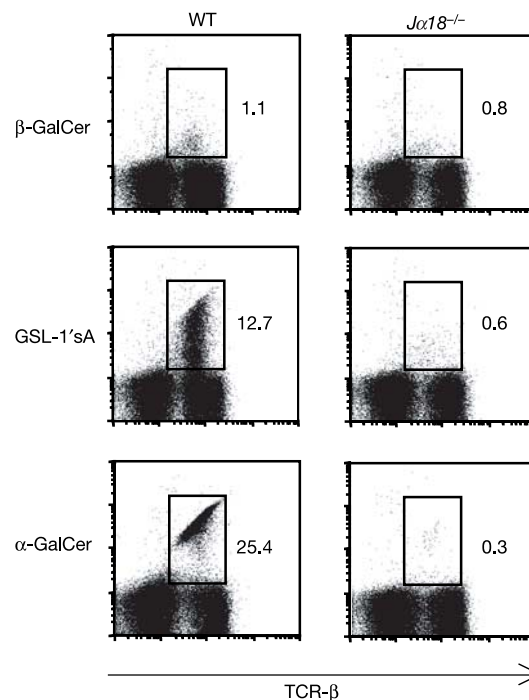
$\alpha$ -GalCer, but this is an agonist of exceptional potency, and the galacturonic-acid-containing GSL-1' was more active than the glucuronic-acid-containing GSL-1, analogous to the increased potency of  $\alpha$ -GalCer containing a galactose sugar compared with the glucose-containing form,  $\alpha$ -glucosyl ceramide<sup>11</sup>.

In addition to having an identical TCR  $\alpha$ -chain, V $\alpha$ 14i NKT cells predominantly express V $\beta$ 8.2, V $\beta$ 7 or V $\beta$ 2  $\beta$ -chains, which are characterized by highly variable CDR3 regions due to extensive diversity generated during TCR rearrangement<sup>12</sup>. It is therefore possible that only a minority of  $\alpha$ -GalCer-reactive V $\alpha$ 14i NKT cells can respond to *Sphingomonas* GSL compounds, because of the expression of TCR  $\beta$ -chains with particular CDR3 regions. An example of this type of subspecificity in  $\alpha$ -GalCer-reactive T cells is illustrated by the relatively small group of V $\alpha$ 14i NKT cells that also respond to ganglioside GD<sub>3</sub> (ref. 13). Therefore, to test for *Sphingomonas* antigen reactivity at the single-cell level in populations of V $\alpha$ 14i NKT cells, we stained liver mononuclear cells (Fig. 2) and spleen cells (data not shown) with CD1d tetramers loaded with different glycolipids, including GSL-1'sA, a synthetic version of the galacturonic acid-containing GSL-1' compound with the 18-carbon sphingosine. GSL-1'sA-loaded CD1d tetramers stained about half as many liver mononuclear cells as the  $\alpha$ -GalCer tetramers, but these cells were absent from *CD1d*<sup>-/-</sup> mice (data not shown), which cannot positively select V $\alpha$ 14i NKT cells. They were also absent from *J $\alpha$ 18*<sup>-/-</sup> mice, which lack V $\alpha$ 14i NKT cells because they do not have the *J $\alpha$*  gene segment required for formation of the V $\alpha$ 14i TCR rearrangement<sup>7</sup>. This demonstrates that T cells binding GSL-1'sA and  $\alpha$ -GalCer use the same TCR rearrangement and are therefore overlapping. Cells that bound the GSL-1'sA-loaded CD1d tetramers had an increased percentage of V $\beta$ 8.2-containing TCRs relative to V $\beta$ 7 TCRs (data not shown). V $\beta$ 8.2-containing TCRs have a higher affinity for  $\alpha$ -GalCer presented by CD1d<sup>14</sup>, and because GSL-1'sA is a less potent agonist than  $\alpha$ -GalCer we would expect these higher-affinity V $\beta$ 8.2 TCRs to be enriched in



**Figure 1** V $\alpha$ 14i NKT-cell hybridomas respond to *Sphingomonas* glycolipids.

**a**, Glycosphingolipid structures. Some of the purified compounds contain a hydroxyl group at the 2 position (designated X) of the acyl chain. Differences from  $\alpha$ -GalCer are shown in red. Synthetic compounds used in this study have the 18-carbon sphingosine and either lack the acyl-chain hydroxyl group (GSL-1'sA) or contain it (GSL-1'sB). **b**, Dose-response curve of V $\alpha$ 14i NKT-cell hybridoma 1.2 (V $\beta$ 8.2/V $\alpha$ 14i) to CD1d-coated plates incubated with the indicated glycolipid concentrations. Release of IL-2 was measured by ELISA after 16 h of culture. Similar results were obtained with hybridoma 1.4 (V $\beta$ 10/V $\alpha$ 14i). **c**, CD1d-reactive but V $\alpha$ 14-negative hybridoma 19 does not respond to CD1d-coated plates containing the *Sphingomonas* antigens but does respond to A20 cells transfected with CD1d. Data in **b** and **c** are means  $\pm$  s.d. for triplicate wells. Representative data from at least three experiments are shown.



**Figure 2** Most V $\alpha$ 14i NKT cells bind GSL-1/CD1d tetramers. Panels show staining of liver mononuclear cells from wild-type (WT) C57BL/6 or *J $\alpha$ 18*<sup>-/-</sup> mice with CD1d tetramers loaded either with  $\beta$ -GalCer,  $\alpha$ -GalCer or synthetic *Sphingomonas* glycosphingolipid GSL-1'sA. Representative data are shown from one of three experiments in which two individual mice were analysed.

cells binding the GSL-1'sA-loaded CD1d tetramers.

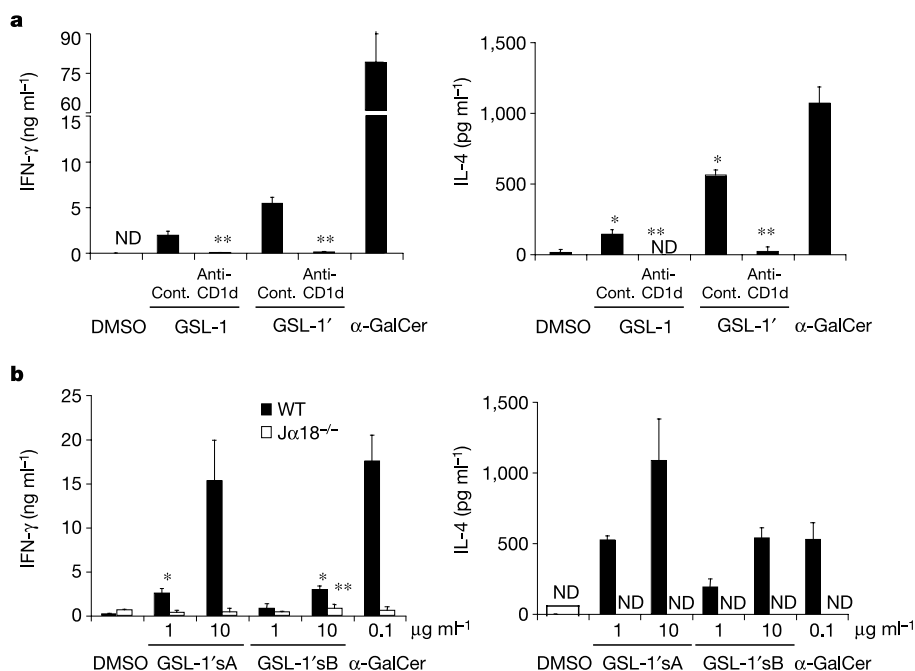
We performed several types of experiments to determine whether V $\alpha$ 14i NKT cells isolated *ex vivo* could respond to *Sphingomonas* GSL compounds when presented by natural antigen-presenting cells (APCs). First, we cultured cells from V $\alpha$ 14i transgenic mice with the purified compounds, then measured cytokine release by enzyme-linked immunosorbent assay (ELISA). These transgenic mice have a roughly fivefold increase in the number of V $\alpha$ 14i NKT cells<sup>15</sup>, thereby permitting optimal measurement of the glycolipid-specific immune response. As shown in Fig. 3a, spleen cells from the transgenic mice produced both IL-4 and interferon (IFN)- $\gamma$  in response to either purified GSL-1 or GSL-1', and this response could be inhibited by anti-CD1d antibodies. As for the V $\alpha$ 14i NKT cell hybridomas (Fig. 1b), the response was stronger to glycolipids containing galacturonic acid than to those containing glucuronic acid. Similar results were obtained with thymocytes (data not shown). Second, we pulsed bone marrow-derived dendritic cells (DCs) with synthetic glycolipids, and cultured these DCs with liver mononuclear cells prepared from either wild-type or *J $\alpha$ 18<sup>-/-</sup>* mice. The *Sphingomonas* glycolipids stimulated the release of IFN- $\gamma$  and IL-4, although GSL-1'sA was more potent than GSL-1'sB. Cells from *J $\alpha$ 18<sup>-/-</sup>* mice did not respond, further indicating that GSL-1 reactivity is dependent on V $\alpha$ 14i NKT cells (Fig. 3b).

To determine whether the *Sphingomonas* glycolipids could stimulate cytokine release and V $\alpha$ 14i NKT cell activation *in vivo*, we immunized mice with bone marrow-derived DCs pulsed with GSL-1'sA or  $\alpha$ -GalCer, and tested cells reactive with CD1d tetramers 12 h later. In agreement with previous results<sup>16</sup>,  $\alpha$ -GalCer-pulsed DCs preferentially stimulated the release of IFN- $\gamma$  by V $\alpha$ 14i NKT cells, although some IL-4 release was also observed (Fig. 4a). When DCs were pulsed with GSL-1'sA, a substantial number of V $\alpha$ 14i NKT cells in the liver were also induced to produce cytokines (Fig. 4a). Not surprisingly, in comparison with the highly potent

agonist  $\alpha$ -GalCer, fewer cells responded to GSL-1'sA, and the mean fluorescence intensity for intracellular cytokine was lower. We have found that the induction of expression of the early activation antigen CD69 is a more sensitive measure of V $\alpha$ 14i NKT activation *in vivo* than is intracellular cytokine staining. Consistent with this is our observation that 12 h after the injection of GSL-1'sA-pulsed DCs, CD69 expression was increased on most of the V $\alpha$ 14i NKT cells (Supplementary Fig. 1). Cytokines could also be detected in the serum after the injection of GSL-1'sA-pulsed DCs (data not shown).

Injection of bacteria or bacteria-pulsed DCs into mice caused the specific *in vivo* activation of V $\alpha$ 14i NKT cells, as measured by the induction of increased expression of CD69, an early activation antigen (Fig. 4b and Supplementary Fig. 1). To determine whether V $\alpha$ 14i NKT cells could influence bacterial clearance *in vivo*, we injected 10<sup>6</sup> bacteria intraperitoneally and measured the number of surviving bacteria 24 or 48 h later. The number of surviving bacteria was significantly increased in the liver of V $\alpha$ 14i NKT-cell-deficient *CD1d<sup>-/-</sup>* mice (Fig. 4c), providing direct evidence for the participation of V $\alpha$ 14i NKT cells in bacterial clearance. Similar results were obtained when a larger dose of bacteria (10<sup>7</sup>) was injected (data not shown). Increased colony counts were also observed with *J $\alpha$ 18<sup>-/-</sup>* mice (data not shown).

Human V $\alpha$ 24i NKT cells do not respond to some  $\alpha$ -GalCer-related glycolipids that can activate mouse V $\alpha$ 14i NKT cells<sup>17</sup>. To determine whether the reactivity for *Sphingomonas* glycolipids is conserved in human V $\alpha$ 24i NKT cells, we stained a human T-cell line, expanded with  $\alpha$ -GalCer, with human CD1d dimers loaded with GSL-1'sA or GSL-1'sB. Essentially all of the V $\alpha$ 24<sup>+</sup> T cells in the line reacted with CD1d dimers loaded with either glycolipid, as well as with  $\alpha$ -GalCer-containing dimers (Fig. 5a). Moreover, the expanded V $\alpha$ 24i NKT cells in the line responded significantly to DCs pulsed with *Sphingomonas* glycolipids by secreting IFN- $\gamma$  (Fig. 5b). The reactivity to the bacterial glycosphingolipids is therefore conserved in human V $\alpha$ 24i NKT cells.



**Figure 3** *In vitro* response of non-transformed V $\alpha$ 14i NKT cells to *Sphingomonas* antigens. **a**, Spleen cells from V $\alpha$ 14i TCR transgenic mice were cultured with dimethyl sulphoxide (DMSO) vehicle or the indicated antigens, and the secretion of cytokines was measured by ELISA after 3 days. ND, not detected. To block reactivity, 20  $\mu$ g ml<sup>-1</sup> anti-mouse CD1d monoclonal antibody 1B1 or the isotype control was added. **b**, Bone marrow-derived DCs were pulsed with the indicated antigens or DMSO for 24 h, and

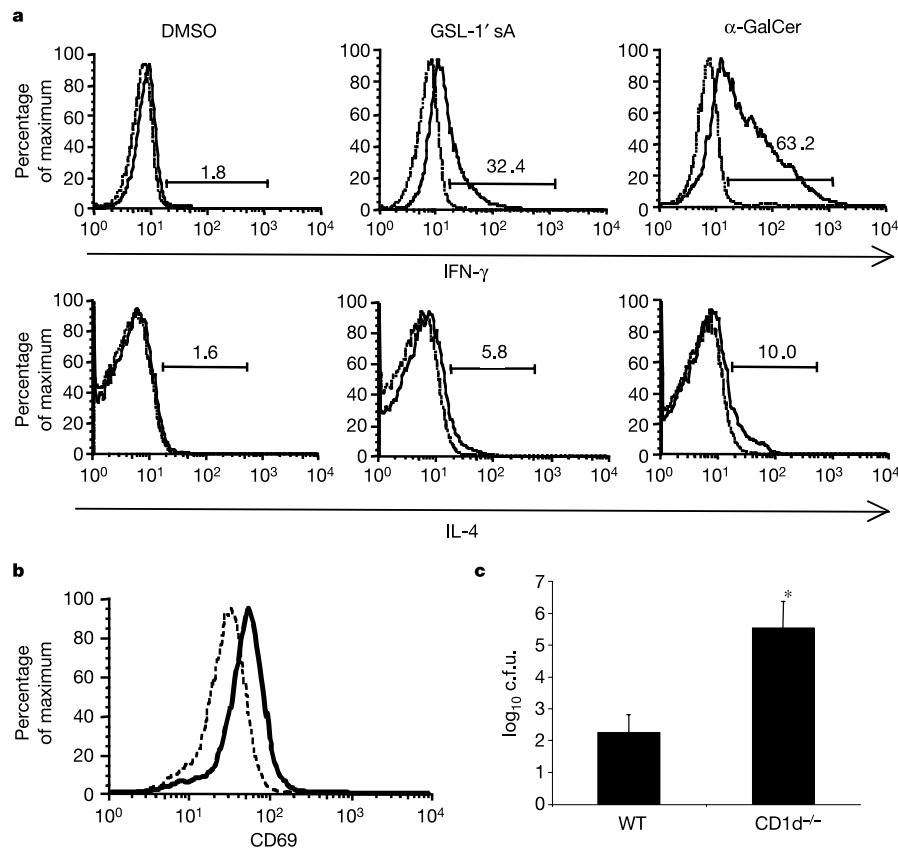
cultured with liver mononuclear cells from wild-type or *J $\alpha$ 18<sup>-/-</sup>* mice. Cytokine secretion in the supernatant was measured after 48 h. Data are means  $\pm$  s.d. for triplicate determinations and were evaluated with Student's *t*-test. Asterisk,  $P < 0.05$ , GSL-1 compared with DMSO. Two asterisks,  $P < 0.05$ , anti-CD1d compared with control IgG (**a**) or *J $\alpha$ 18<sup>-/-</sup>* mice compared with wild-type mice (**b**). Representative data from at least three experiments are shown in each case.



There are several infections in which V $\alpha$ 14i NKT cells become activated and expanded and where they might be essential for host defence<sup>18,19</sup>, but it is much less certain whether microbial antigens were being recognized in these cases. V $\alpha$ 14i NKT-cell responses to inflammatory signals in the absence of CD1d expression have been observed<sup>20</sup> and it has been reported that the activation of V $\alpha$ 14i and V $\alpha$ 24i NKT cells in response to infection with *Salmonella* does not involve microbial antigen recognition<sup>21</sup>. Instead, activation of Toll-like receptors (TLRs) by lipopolysaccharide (LPS) from *Salmonella* was reported to induce the production of IL-12 by APCs, and this IL-12, combined with recognition of self antigens presented by CD1d, was sufficient to activate NKT cells in the absence of any microbial antigen<sup>21</sup>. To determine whether *Sphingomonas* GSL-1 could activate V $\alpha$ 14i NKT cells indirectly by stimulating APCs, we cultured DCs with GSL-1'sA and tested them for maturation (data not shown) and cytokine release (Supplementary Fig. 2). Culture with GSL-1'sA did not activate DCs. Furthermore, the response to *Sphingomonas* is not dependent on TLR4, because the activation of V $\alpha$ 14i NKT cells by *Sphingomonas* bacteria did not depend on TLR4 expression (Supplementary Fig. 3) and bacterial clearance was not decreased in TLR4-deficient C3H/HeJ mice (data not shown). DCs from mice with mutations in *MyD88* and *Trif*, which lack all TLR signalling, could still activate V $\alpha$ 14i NKT cells to produce IFN- $\gamma$  (Supplementary Fig. 4), as could DCs from *IL-12*<sup>-/-</sup> mice (Supplementary Figs 5 and 6). These experiments demonstrate that *Sphingomonas* GSL-1'sA acts similarly to  $\alpha$ -GalCer, primarily by engaging the TCR expressed by V $\alpha$ i NKT cells.

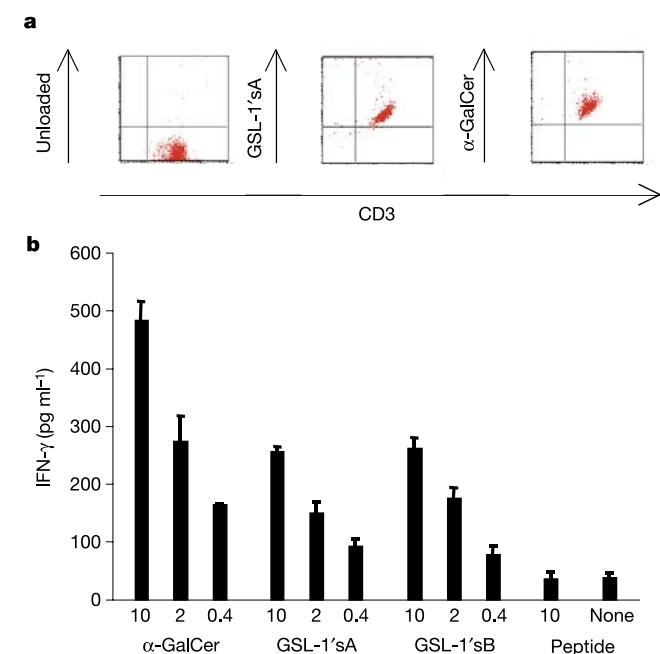
Although the concept of direct microbial antigen reactivity mediated by the TCRs of mouse and human V $\alpha$ i NKT cells has been controversial, it was reported recently that relatively small proportions of these cells react to a glycolipid from mycobacterial cell walls, phosphatidylinositol tetramannoside<sup>22</sup>, or to a lipophosphoglycan from *Leishmania donovani*<sup>23</sup>. Because only a small minority of the V $\alpha$ 14i and V $\alpha$ 24i NKT cells responded, these specificities are presumably related to the expression of particular TCR  $\beta$ -chains, perhaps in a manner similar to that of the GD3-reactive T cells mentioned above. These antigens therefore cannot account for the selection and conservation of an expanded, activated population of cells with the V $\alpha$ 14i and V $\alpha$ 24i rearrangements. However, our present work shows that some microbes contain antigenic glycosphingolipids with  $\alpha$ -linked sugars related to  $\alpha$ -GalCer that can stimulate most V $\alpha$ 14i and V $\alpha$ 24i NKT cells, indicating that compounds with this ability are more widely distributed than previously thought, rather than being confined to marine organisms.

V $\alpha$ 14i NKT cells are relatively abundant in some tissues; they have an antigen-experienced phenotype and can rapidly produce cytokines such as IL-4 and IFN- $\gamma$  even in the absence of prior exposure to exogenous antigens<sup>24</sup>. Because they respond rapidly and do not require clonal expansion, the kinetics and magnitude of their response are similar in some ways to innate immune responses. It is therefore not surprising that the activation of V $\alpha$ 14i NKT cells by  $\alpha$ -GalCer provides an adjuvant-like stimulus for conventional T-cell responses<sup>25,26</sup>, similar to that of LPS and other microbial substances. *Sphingomonas* cells lack LPS, and therefore perhaps in this and other



**Figure 4** *In vivo* response of V $\alpha$ 14i NKT cells to bacterial antigens and bacteria. **a**, Intracellular cytokine staining of gated,  $\alpha$ -GalCer CD1d tetramer-positive liver mononuclear cells 12 h after intravenous injection of DCs pulsed with vehicle or the indicated glycolipid. Representative data are shown from three experiments. DMSO, dimethyl sulfoxide. **b**, Induction of CD69 expression by gated,  $\alpha$ -GalCer CD1d tetramer-positive liver mononuclear cells 12 h after intravenous injection of DCs pulsed with live

*S. yanoikuyae* (solid line) or unpulsed DCs (dotted line). Representative data are shown from one of two experiments. **c**, Surviving bacteria in the livers of wild-type (WT) and *CD1d*<sup>-/-</sup> mice 48 h after injection of  $10^6$  *S. yanoikuyae*. Data are means  $\pm$  s.d. for five mice in each group and were evaluated with Student's *t*-test. Asterisk *P* < 0.05. Representative data from one of three experiments are shown.



**Figure 5** Human Vα24i NKT cells respond to synthetic *Sphingomonas* glycolipids. **a**, Staining of a human Vα24i NKT-cell line with human CD1d dimers that were unloaded or loaded with the indicated glycosphingolipids. GSL-1'sB-loaded dimers stained equivalently to GSL-1'sA-containing dimers (data not shown). **b**, IFN-γ release, after 16 h of culture, by a human Vα24i NKT cell line in response to culture with autologous DCs pulsed with the indicated glycolipid antigen concentrations (μg ml⁻¹). Data are means ± s.d. for triplicate wells. Representative data from one of three experiments are shown.

cases a rearranged and highly selected TCR can serve as a surrogate pattern recognition receptor for the recognition of bacterial glycosphingolipids presented by CD1d. As many as 30% of the lymphocytes in the liver are Vα14i NKT cells, so we consider it highly likely that the TCR-mediated activation of IFN-γ release by most of these cells is responsible for the effect of Vα14i NKT cells on *Sphingomonas* growth in the liver. However, it remains formally possible that some other means of activating these cells *in vivo* is equally or even more important for increased clearance of *Sphingomonas* in mice that have Vα14i NKT cells.

In recent years the concept has emerged that specialized populations of lymphocytes are inherently self-reactive<sup>27</sup>. In some cases, these lymphocyte subpopulations have a highly restricted antigen receptor diversity, and the self specificities they exhibit might therefore be quite limited. B1 B cells provide one such example of restricted diversity and apparently beneficial self-reactivity for phosphocholine. Although phosphocholine-specific antibodies are natural antibodies that react with apoptotic cells<sup>28</sup> and therefore might be involved in cell clearance, this specificity might also have been selected for in evolution because of its reactivity for microbial pathogens. We speculate that Vαi NKT cells might be similar, in that their self-reactivity and oligoclonality were selected in evolution by a beneficial self-reactivity for iGb3 and perhaps related compounds that could also serve the dual role of providing protection against some pathogens.

## Methods

### Reagents and mice

α-GalCer and β-GalCer were obtained from the Kirin pharmaceutical research corporation. The *Sphingomonas* glycolipids were purified as described previously<sup>3–5</sup>, with the GSL-1 used in these experiments from *S. paucimobilis* and the GSL-1' from *S. yanoikuyae*. GSL-1 from *S. yanoikuyae* is also antigenic but contains a small amount of GSL-1'; we therefore used the *S. paucimobilis* material for most experiments. Evidence for

an α-anomeric configuration of the sugars has been presented<sup>5</sup>, and additional chemical details will be published (K.K., manuscript in preparation). The formation of synthetic (s) compounds, including GSL-1'sA and GSL-1'sB, will be described in detail (D.W., G.-W.X. and C.-H.W., manuscript in preparation). CD1d tetramers loaded with different glycolipids were prepared and used to stain cells as described<sup>29</sup>. CD1d-reactive hybridomas have been described previously<sup>17,29</sup>. Vα14i transgenic mice, Jα18<sup>-/-</sup> mice and CD1d<sup>-/-</sup> mice on the C57BL/6 background were gifts from A. Bendelac, M. Taniguchi and L. van Kaer, respectively. All mice were housed under specific pathogen-free conditions and the experiments were approved by the Institutional Animal Care and Use Committee of the La Jolla Institute of Allergy and Immunology.

### Immune assays

Stimulation of T-cell hybridomas on CD1d-coated plates and intracellular cytokine staining of tetramer-positive cells were performed as described in published protocols<sup>11,30</sup>. Mouse DCs were prepared by culturing bone marrow progenitors with mouse granulocyte/macrophage colony-stimulating factor (PeproTech) for 6 days and were incubated with the indicated concentrations of glycolipids and washed, and 4 × 10<sup>4</sup> DCs were cultured with 2 × 10<sup>5</sup> liver mononuclear cells. Supernatants were harvested and cytokines were measured by ELISA. Alternatively, 5 × 10<sup>5</sup> glycolipid-pulsed DCs were injected intravenously and α-GalCer/CD1d-tetramer-positive cells were analysed 12 h later for intracellular cytokines. *S. yanoikuyae* cells were grown in tryptic soy broth at 30 °C, and 10<sup>6</sup> or 10<sup>7</sup> bacteria were injected intraperitoneally or 5 × 10<sup>7</sup> bacteria were added to 10<sup>6</sup> DCs, before the intravenous injection of 5 × 10<sup>5</sup> DCs. Induction of CD69 expression by Vα14i NKT cells was measured 12 h after the injection of bacteria-pulsed DCs or 24 h after the injection of bacteria. Bacterial counts were performed by plating homogenized liver tissue (harvested 24 or 48 h after bacterial injection) on tryptic soy agar plates, and individual colonies were counted.

### Generation and analysis of human natural killer cell lines

Immature DCs were generated from CD14<sup>+</sup> cells in leukopaks after a two-day incubation in the presence of 300 U ml<sup>-1</sup> granulocyte/macrophage colony-stimulating factor (R&D Systems) and 100 U ml<sup>-1</sup> IL-4 (R&D Systems). After irradiation (3,000 rad), the immature DCs were cultured together with syngeneic CD161<sup>+</sup> cells in the presence of 100 ng ml<sup>-1</sup> α-GalCer and 10 U ml<sup>-1</sup> IL-2 for 10 days. After repeating this stimulation, two natural killer cell lines were shown to express both CD161 and a Vα24i TCR (99% purity). Soluble divalent human CD1d-IgG<sub>1</sub> fusion protein (1 μg; BD Pharmingen) was incubated overnight with glycolipids in accordance with the manufacturer's protocol. The glycolipid-loaded CD1d-IgG<sub>1</sub> dimers were incubated with human NKT cells at 4 °C for 60 min, followed by incubation with phycoerythrin-coupled anti-mouse IgG<sub>1</sub> monoclonal antibody (A85-1). Human natural killer cells (5 × 10<sup>4</sup>) and 2 × 10<sup>4</sup> irradiated immature autologous DCs were cultured together in the presence of the glycolipid compounds at the indicated concentrations in a 96-well plate. After culture for 16 h, the concentration of IFN-γ in the culture supernatants was determined by ELISA (BD Pharmingen).

Received 12 December 2004; accepted 25 January 2005; doi:10.1038/nature03407.

1. Taniguchi, M., Harada, M., Kojo, S., Nakayama, T. & Wakao, H. The regulatory role of Vα14 NKT cells in innate and acquired immune response. *Annu. Rev. Immunol.* **21**, 483–513 (2003).
2. Brigl, M. & Brenner, M. B. CD1: antigen presentation and T cell function. *Annu. Rev. Immunol.* **22**, 817–890 (2004).
3. Kawahara, K. *et al.* Chemical structure of glycosphingolipids isolated from *Sphingomonas paucimobilis*. *FEBS Lett.* **292**, 107–110 (1991).
4. Kawahara, K., Moll, H., Knirel, Y. A., Seydel, U. & Zahring, U. Structural analysis of two glycosphingolipids from the lipopolysaccharide-lacking bacterium *Sphingomonas capsulata*. *Eur. J. Biochem.* **267**, 1837–1846 (2000).
5. Kawahara, K., Kubota, M., Sato, N., Tsuge, K. & Seto, Y. Occurrence of an alpha-galacturonosylceramide in the dioxin-degrading bacterium *Sphingomonas wittichii*. *FEMS Microbiol. Lett.* **214**, 289–294 (2002).
6. Kobayashi, E., Motoki, K., Uchida, T., Fukushima, H. & Koezuka, Y. KRN7000, a novel immunomodulator, and its antitumor activities. *Oncol. Res.* **7**, 529–534 (1995).
7. Kawano, T. *et al.* CD1d-restricted and TCR-mediated activation of Vα14 NKT cells by glycosylceramides. *Science* **278**, 1626–1629 (1997).
8. Brossay, L. *et al.* CD1d-mediated recognition of an α-galactosylceramide by natural killer T cells is highly conserved through mammalian evolution. *J. Exp. Med.* **188**, 1521–1528 (1998).
9. Zhou, D. *et al.* Lysosomal glycosphingolipid recognition by NKT cells. *Science* **306**, 1786–1789 (2004).
10. Neef, A., Witzemberger, R. & Kampfer, P. Detection of sphingomonads and *in situ* identification in activated sludge using 16S rRNA-targeted oligonucleotide probes. *J. Ind. Microbiol. Biotechnol.* **23**, 261–267 (1999).
11. Sidobre, S. *et al.* The T cell antigen receptor expressed by Vα14i NKT cells has a unique mode of glycosphingolipid antigen recognition. *Proc. Natl Acad. Sci. USA* **101**, 12254–12259 (2004).
12. Matsuda, J. L. *et al.* Natural killer T cells reactive to a single glycolipid exhibit a highly diverse T cell receptor β repertoire and small clone size. *Proc. Natl Acad. Sci. USA* **98**, 12636–12641 (2001).
13. Wu, D. Y., Segal, N. H., Sidobre, S., Kronenberg, M. & Chapman, P. B. Cross-presentation of disialoganglioside GD3 to natural killer T cells. *J. Exp. Med.* **198**, 173–181 (2003).
14. Schumann, J., Voyle, R. B., Wei, B. Y. & MacDonald, H. R. Cutting edge: influence of the TCR V β domain on the avidity of CD1d:α-galactosylceramide binding by invariant Vα14 NKT cells. *J. Immunol.* **170**, 5815–5819 (2003).
15. Lehen, A. *et al.* Overexpression of natural killer T cells protects Vα14- Jα281 transgenic nonobese diabetic mice against diabetes. *J. Exp. Med.* **188**, 1831–1839 (1998).
16. Fujii, S., Shimizu, K., Kronenberg, M. & Steinman, R. M. Prolonged IFN-γ-producing NKT response induced with α-galactosylceramide-loaded DCs. *Nature Immunol.* **3**, 867–874 (2002).
17. Brossay, L. *et al.* Structural requirements for galactosylceramide recognition by CD1-restricted NK T cells. *J. Immunol.* **161**, 5124–5128 (1998).

18. Vincent, M. S., Gumperz, J. E. & Brenner, M. B. Understanding the function of CD1-restricted T cells. *Nature Immunol.* **4**, 517–523 (2003).
19. Skold, M. & Behar, S. M. Role of CD1d-restricted NKT cells in microbial immunity. *Infect. Immun.* **71**, 5447–5455 (2003).
20. Mempel, M. *et al.* Natural killer T cells restricted by the monomorphic MHC class 1b CD1d molecules behave like inflammatory cells. *J. Immunol.* **168**, 365–371 (2002).
21. Brigl, M., Bry, L., Kent, S. C., Gumperz, J. E. & Brenner, M. B. Mechanism of CD1d-restricted natural killer T cell activation during microbial infection. *Nature Immunol.* **4**, 1230–1237 (2003).
22. Fischer, K. *et al.* Mycobacterial phosphatidylinositol mannoside is a natural antigen for CD1d-restricted T cells. *Proc. Natl Acad. Sci. USA* **101**, 10685–10690 (2004).
23. Amprey, J. L. *et al.* A subset of liver NK T cells is activated during *Leishmania donovani* infection by CD1d-bound lipophosphoglycan. *J. Exp. Med.* **200**, 895–904 (2004).
24. Park, S. H., Benlagha, K., Lee, D., Balish, E. & Bendelac, A. Unaltered phenotype, tissue distribution and function of V $\alpha$ 14<sup>+</sup> NKT cells in germ-free mice. *Eur. J. Immunol.* **30**, 620–625 (2000).
25. Gonzalez-Aseguinolaza, G. *et al.* Natural killer T cell ligand  $\alpha$ -galactosylceramide enhances protective immunity induced by malaria vaccines. *J. Exp. Med.* **195**, 617–624 (2002).
26. Fujii, S., Shimizu, K., Smith, C., Bonifaz, L. & Steinman, R. M. Activation of natural killer T cells by  $\alpha$ -galactosylceramide rapidly induces the full maturation of dendritic cells *in vivo* and thereby acts as an adjuvant for combined CD4 and CD8 T cell immunity to a coadministered protein. *J. Exp. Med.* **198**, 267–279 (2003).
27. Cheroutre, H. Starting at the beginning: new perspectives on the biology of mucosal T cells. *Annu. Rev. Immunol.* **22**, 217–246 (2004).
28. Shaw, P. X., Goodyear, C. S., Chang, M. K., Witztum, J. L. & Silverman, G. J. The autoreactivity of anti-phosphorylcholine antibodies for atherosclerosis-associated neo-antigens and apoptotic cells. *J. Immunol.* **170**, 6151–6157 (2003).
29. Matsuda, J. L. *et al.* Tracking the response of natural killer T cells to a glycolipid antigen using CD1d tetramers. *J. Exp. Med.* **192**, 741–754 (2000).
30. Matsuda, J. L. *et al.* Mouse V $\alpha$ 14i natural killer T cells are resistant to cytokine polarization *in vivo*. *Proc. Natl Acad. Sci. USA* **100**, 8395–8400 (2003).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank K. J. L. Hammond for critical reading of the manuscript, and S. Sidobre, L. Sidobre, K. J. L. Hammond and A. Khurana for mCD1d protein. This work was supported by grants from the National Institutes of Health (to M.K., to C.-H.W. and to M.T.). Y.K. was supported in part by the Yamada Science Foundation.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to M.K. ([mitch@liai.org](mailto:mitch@liai.org)).

## Exogenous and endogenous glycolipid antigens activate NKT cells during microbial infections

Jochen Mattner<sup>1</sup>, Kristin L. DeBord<sup>2</sup>, Nahed Ismail<sup>3</sup>, Randal D. Goff<sup>4</sup>, Carlos Cantu III<sup>5</sup>, Dapeng Zhou<sup>1</sup>, Pierre Saint-Mezard<sup>1</sup>, Vivien Wang<sup>1</sup>, Ying Gao<sup>4</sup>, Ning Yin<sup>4</sup>, Kasper Hoebe<sup>5</sup>, Olaf Schneewind<sup>2</sup>, David Walker<sup>3</sup>, Bruce Beutler<sup>5</sup>, Luc Teyton<sup>5</sup>, Paul B. Savage<sup>4\*</sup> & Albert Bendelac<sup>1\*</sup>

<sup>1</sup>Committee on Immunology and <sup>2</sup>Committee on Microbiology, University of Chicago, Chicago, Illinois 60637, USA

<sup>3</sup>University of Texas Medical Branch, Department of Pathology, Galveston, Texas 77555, USA

<sup>4</sup>Brigham Young University, Department of Chemistry and Biochemistry, Provo, Utah 84602-5700, USA

<sup>5</sup>The Scripps Research Institute, Department of Immunology, La Jolla, California 92037, USA

\* These authors contributed equally to this work

CD1d-restricted natural killer T (NKT) cells are innate-like lymphocytes that express a conserved T-cell receptor and contribute to host defence against various microbial pathogens<sup>1,2</sup>. However, their target lipid antigens have remained elusive. Here we report evidence for microbial, antigen-specific activation of NKT cells against Gram-negative, lipopolysaccharide (LPS)-negative alpha-Proteobacteria such as *Ehrlichia muris* and *Sphingomonas capsulata*. We have identified glycosylceramides

from the cell wall of *Sphingomonas* that serve as direct targets for mouse and human NKT cells, controlling both septic shock reaction and bacterial clearance in infected mice. In contrast, Gram-negative, LPS-positive *Salmonella typhimurium* activates NKT cells through the recognition of an endogenous lysosomal glycosphingolipid, iGb3, presented by LPS-activated dendritic cells. These findings identify two novel antigenic targets of NKT cells in antimicrobial defence, and show that glycosylceramides are an alternative to LPS for innate recognition of the Gram-negative, LPS-negative bacterial cell wall.

CD1 is a family of  $\beta$ 2-microglobulin-associated, major histocompatibility complex-like molecules that evolved to capture lipid antigens in different cellular compartments for display at the surface of antigen-presenting cells<sup>1,2</sup>. A variety of self and microbial lipid antigens can be presented by CD1a, b and c for specific recognition by  $\alpha\beta$  T cells expressing diverse T-cell receptors (TCRs). In contrast, CD1d is associated with an innate-like population of memory/effector, NK receptor-expressing NKT cells that predominantly use a conserved, semi-invariant mouse V $\alpha$ 14-J $\alpha$ 18/V $\beta$ 8 or human V $\alpha$ 24-J $\alpha$ 18/V $\beta$ 11 TCR, suggesting different modalities of activation. Like NK cells, NKT cells constitutively express messenger RNA but not protein for interferon- $\gamma$  (IFN- $\gamma$ ), a hallmark of their poised effector stage<sup>3</sup>. NKT-deficient mice have impaired antimicrobial defence due in part to defective early IFN- $\gamma$  secretion<sup>4–6</sup>, but the mechanism of NKT cell activation during infection is only partially understood. *In vitro*, dendritic cells (DCs) pre-treated with *Salmonella typhimurium* extract induce the production of IFN- $\gamma$  by NKT cells in a CD1d-dependent manner, indicating a requirement for CD1d-mediated presentation of antigen. DCs and high doses of semi-purified LPS or interleukin-12 (IL-12) also induced the production of IFN- $\gamma$  by NKT cells in a CD1d-dependent manner, suggesting the unusual possibility that self-antigens might in fact serve as targets of NKT cells in some conditions. However, because studies so far have failed to identify the putative lipid targets of NKT cells, with the exception of phosphatidylinositolmannosides expressed by mycobacteria<sup>7</sup>, the mechanisms of their *in vivo* recruitment during infection remain a matter of speculation.

In a co-culture system combining fresh, bone marrow-derived DCs (BMDCs) and CD1d tetramer-sorted NKT cells, the addition of various heat-killed bacteria induced substantial IFN- $\gamma$  secretion by mouse NKT cells in a CD1d-dependent manner (Fig. 1a, left). Similar results were found with a human NKT cell line co-cultured with DCs derived from peripheral blood mononuclear cells (PBMCs), including the requirement for CD1d expression shown by antibody blocking (Fig. 1a, right). Whole spleen cell suspensions cultured in the presence of these heat-killed bacteria for six days showed a marked expansion and proliferation of NKT cells, only slightly inferior to that induced by pure  $\alpha$ -galactosylceramide ( $\alpha$ GalCer) KR7000, a pharmacological NKT cell ligand (Fig. 1b and Supplementary Fig. S1). These bacteria included Gram-negative, LPS-positive *Salmonella typhimurium* and also, surprisingly, the Gram-negative *Ehrlichia muris*, a lethal pathogen which, unlike *Salmonella*, does not express LPS<sup>8</sup>, as well as *Sphingomonas capsulata*, another Gram-negative LPS-negative member of the same class of alpha-Proteobacteria whose cell wall lipids have been extensively characterized<sup>9</sup> (Fig. 1). We tested the involvement of Toll-like receptors (TLRs) in these anti-microbial responses by examining *MyD88*<sup>−/−</sup>, *Trif*<sup>Δps2/Δps2</sup> and *MyD88*<sup>−/−</sup> *Trif*<sup>Δps2/Δps2</sup> spleen cells lacking one or both of the adaptors, MyD88 and Trif (also known as Ticam-1), necessary for TLR signalling<sup>10</sup>. In the whole spleen cell culture assay, *Salmonella*-induced IFN- $\gamma$  was drastically reduced to 2–15% of control, on average, in the absence of either one or both TLR adaptors (Supplementary Fig. S2). In sharp contrast, the splenic IFN- $\gamma$  response to LPS-negative *Ehrlichia* and *Sphingomonas* was largely independent of MyD88 and Trif. *CD1d*<sup>−/−</sup> spleen cells lacking NKT cells failed to respond to

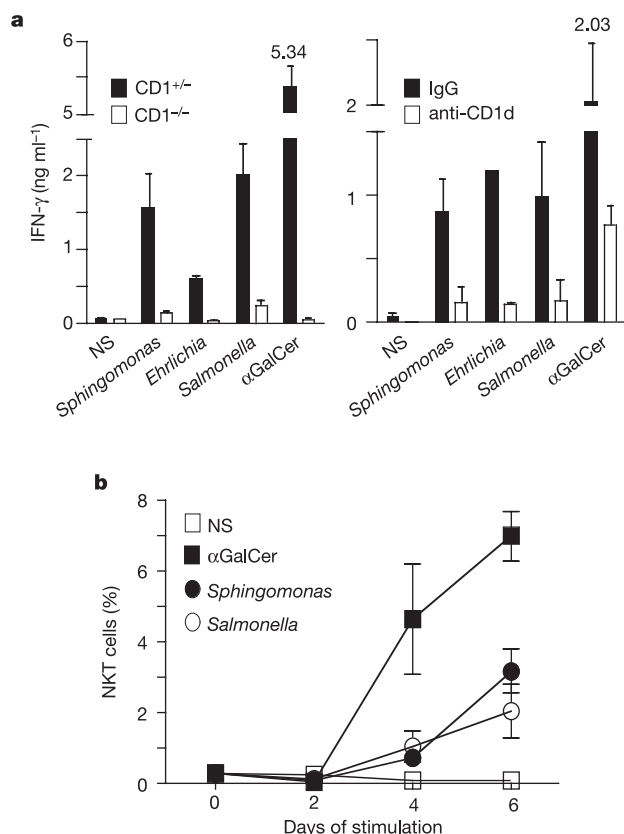


*Sphingomonas* and *Ehrlichia*, whereas the response to *Salmonella* was only marginally reduced. Similarly, wild-type NKT cells co-cultured with MyD88-deficient DCs responded to *Sphingomonas* and *Ehrlichia* but not *Salmonella* (Fig. 2a). These results suggest that in total spleens exposed to heat-killed *Salmonella*, IFN- $\gamma$  production is initiated after TLR signalling of antigen-presenting cells and subsequent recruitment of NKT cells as well as other cell types such as NK cells. In contrast, IFN- $\gamma$  stimulation by *Ehrlichia* and *Sphingomonas* was primarily dependent on NKT cells and CD1d, with minimal contribution from TLRs.

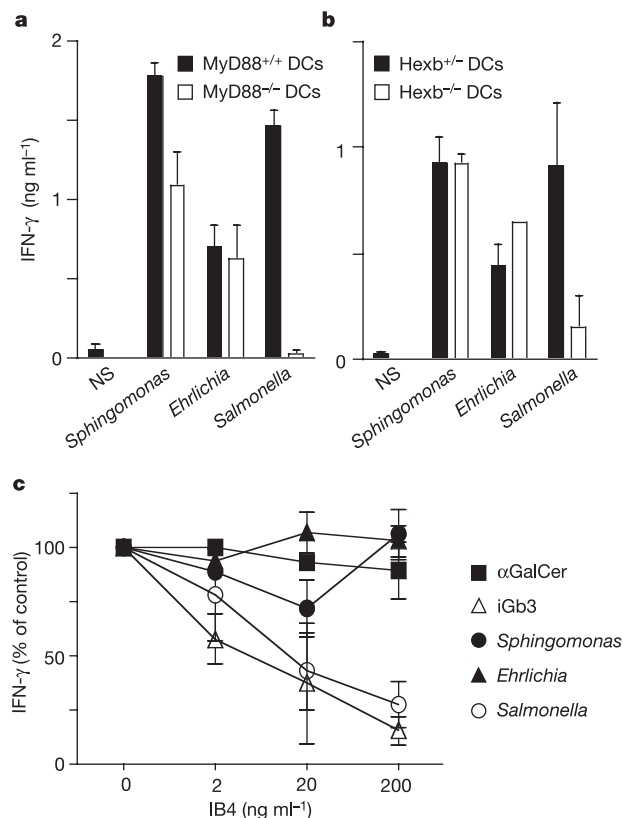
We have recently identified the glycosphingolipid isoglobotrihexosylceramide (iGb3) as a major species of endogenous glycolipid antigens recognized by NKT cells in healthy, non-infected cells<sup>11</sup>. *Hexb*<sup>-/-</sup> DCs fail to generate iGb3 in the lysosome because they lack the  $\beta$ -hexosaminidase required to remove the terminal GalNAc of iGb3, the precursor to iGb3 (ref. 12). *Hexb*<sup>-/-</sup> DCs pulsed with heat-killed *Ehrlichia* or *Sphingomonas* were able to stimulate NKT cells as well as wild-type DCs. In contrast, *Salmonella*-pulsed *Hexb*<sup>-/-</sup> DCs did not stimulate NKT cells (Fig. 2b). The terminal Gal( $\alpha$ 1-3)Gal disaccharide of iGb3 is specifically recognized by the lectin IB4<sup>13</sup>. This lectin also blocked the recognition of iGb3 but not  $\alpha$ GalCer presented by CD1d to NKT cells (Fig. 2c), thus serving as a probe for endogenous ligand recognition by NKT cells. The lectin did not impair the stimulation

of NKT cells by DCs pulsed with heat-killed *Ehrlichia* or *Sphingomonas*, consistent with direct recognition of a distinct microbial antigen. However, the lectin readily blocked stimulation by *Salmonella* (Fig. 2c). Together, the role of  $\beta$ -hexosaminidase and the blockade by Gal( $\alpha$ 1-3)Gal-specific lectins define a sequence of three carbohydrates,  $\beta$ -hexosamine/Gal( $\alpha$ 1-3)/Gal, which is highly characteristic of the isoglobos-series of endogenous glycolipids. Combined with the report that IL-12 or LPS can replace *Salmonella* in the DC-NKT co-culture system<sup>6</sup>, our results identify the endogenous ligand iGb3 rather than a microbial antigen as the target of NKT cells in their response to *Salmonella* infection.

*Ehrlichia* and *Sphingomonas* are examples of Gram-negative bacteria that do not express LPS<sup>8,9,14</sup>, and belong to the same class of alpha-Proteobacteria. Recent analysis of the *Sphingomonas* cell wall revealed the presence of a family of monoglycosylceramides that are substitutes for LPS<sup>9</sup>. Because of their structural analogies with KRN7000, a pharmacological ligand of NKT cells, we synthesized several of the *Sphingomonas* cell wall glycosylceramides and tested their immunological properties. Both  $\alpha$ -glucuronosylceramide (PBS 30) and to a lesser degree  $\alpha$ -galacturonosylceramide (PBS 59) (Fig. 3a) strongly activated mouse and human NKT cell proliferation as well as IFN- $\gamma$  secretion, whereas control  $\beta$ -glucuronosylceramide (PBS 50) did not (Fig. 3b and data not shown). Tetramers of CD1d- $\alpha$ -glucuronosylceramide stained all human NKT cells and ~25% of mouse NKT cells (Fig. 3c). In clonal assays, however, 16/16 different mouse V $\alpha$ 14 hybridomas responded to both  $\alpha$ -glucuronosylceramide and  $\alpha$ -galacturonosylceramide, suggesting that tetramer staining underestimated the



**Figure 1** CD1d-restricted antimicrobial NKT cell responses. **a**, Left, purified mouse V $\alpha$ 14 NKT cells and CD1<sup>+/+</sup> or CD1<sup>-/-</sup> BMDCs were co-cultured in the presence of  $5 \times 10^6$  heat-killed bacteria or 100 ng ml<sup>-1</sup>  $\alpha$ GalCer as indicated (NS, no bacteria), and IFN- $\gamma$  release was measured after 48 h. Right, similar experiment with a human V $\alpha$ 24 NKT cell line and PBMC-derived DCs, in the presence of 1  $\mu$ g ml<sup>-1</sup> anti-CD1d or control IgG1. **b**, Whole spleen cells from B6 mice were stimulated for 6 days with  $5 \times 10^6$  heat-killed bacteria or 100 ng ml<sup>-1</sup>  $\alpha$ GalCer, and the frequency of CD1d- $\alpha$ GalCer<sup>+</sup> NKT cells measured at indicated time points. Absolute numbers of lymphocytes were comparable for each stimulus and time point analysed. Data in **a** and **b** show means  $\pm$  s.d. from three separate experiments.

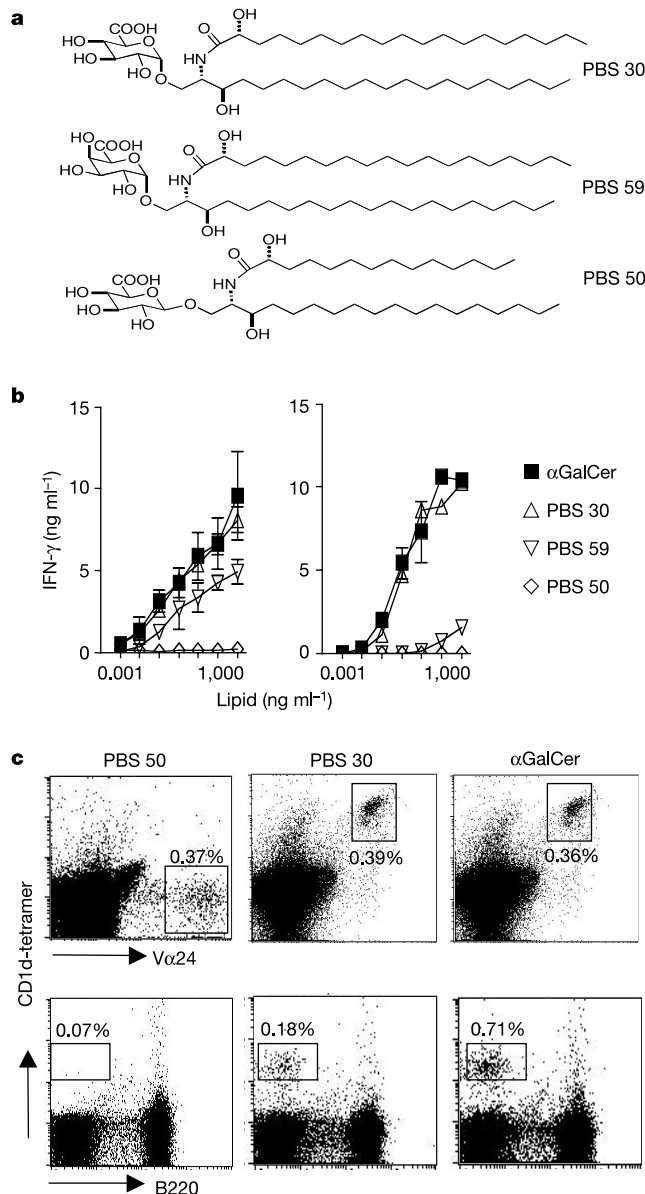


**Figure 2** Differential requirements for the IFN- $\gamma$  response to *Sphingomonas* and *Ehrlichia* versus *Salmonella*. **a**, **b**, IFN- $\gamma$  released by purified mouse NKT cells cultured with MyD88<sup>+/+</sup> or MyD88<sup>-/-</sup> (**a**), Hexb<sup>+/+</sup> or Hexb<sup>-/-</sup> DCs (**b**) and heat-killed bacteria. Data show means  $\pm$  s.d. from two or three separate experiments. **c**, Blockade of human NKT cell responses to DCs + antigen by the lectin IB4, which is specific for the [Gal( $\alpha$ 1-3)Gal] epitope of iGb3 bound to CD1d. Points show means  $\pm$  s.d. from two experiments.

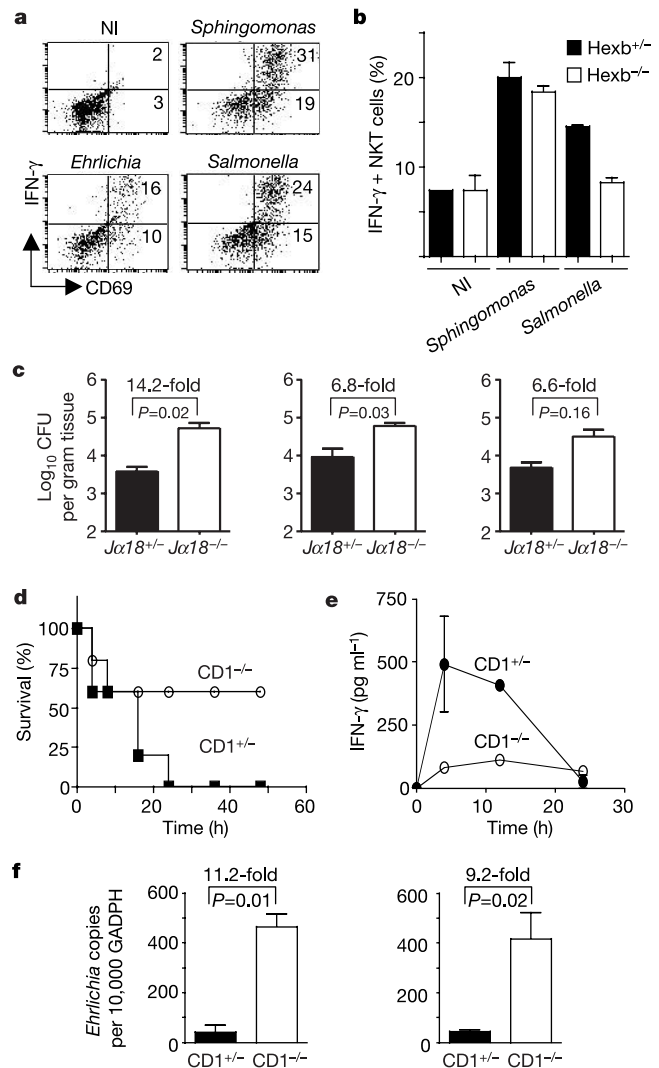
frequency of responders (not shown). These findings reveal that the lipids replacing LPS in the cell wall of at least some Gram-negative bacteria have themselves become targets of innate immunity, as they are directly recognized by the conserved TCR of innate-like NKT cells.

We conclude therefore that the Gram-negative, LPS-positive bacterium *Salmonella* acts primarily through TLR activation of antigen-presenting cells, subsequently inducing NKT cells to produce IFN- $\gamma$  through the combined stimulation by their endogenous iGb3 antigen and IL-12. NKT cells were just one of several IFN- $\gamma$ -producing cell-types recruited by *Salmonella*, explaining why NKT-deficient mice do not appear to be particularly susceptible to *Salmonella* (data not shown). In contrast, Gram-

negative, LPS-negative *Ehrlichia* and *Spingomonas* activate NKT cells primarily through direct recognition of microbial lipids, which include but are not necessarily restricted to cell-wall glycosylceramides (in the case of *Spingomonas*). These microbial



**Figure 3** Different lipid antigens stimulate NKT cells during microbial infections. **a**, Structures of synthetic *Spingomonas* cell wall antigens PBS 30 and PBS 59. PBS 50 is a control  $\beta$ -glucuronosylceramide. **b**, IFN- $\gamma$  response of a human V $\alpha$ 24-J $\alpha$ 18 NKT line (left) and fresh purified mouse NKT cells (right) stimulated by the indicated lipid antigens and DCs. Points shown means  $\pm$  s.d. from two experiments. **c**, CD1d-tetramer staining of fresh human NKT cells (upper row) and fresh mouse spleen cells (lower row) with indicated lipid antigens. NKT cell gate and percentage as indicated. Similar staining results were obtained for fresh or cultured NKT cells from three individuals.



**Figure 4** *In vivo* role of NKT cells during microbial infection. **a**, *In vivo* activation of NKT cells 24 h after intravenous infection with *Spingomonas* ( $1 \times 10^7$ ), *Ehrlichia* ( $1 \times 10^8$ ) and *Salmonella* ( $1 \times 10^6$ ). NI, not infected. NKT cells gated as tetramer $^{+}$  B220 $^{-}$  cells were analysed for surface CD69 and intracellular IFN- $\gamma$ . Similar results were obtained in two experiments. **b**, IFN- $\gamma$  production by NKT cells in response to *Salmonella* requires a Hexb-sufficient host. CFSE-labelled V $\alpha$ 14 transgenic thymocytes were injected intrasplenically 2 h after intraperitoneal challenge with  $5 \times 10^6$  *Spingomonas* or *Salmonella*. Intracellular staining for IFN- $\gamma$  was performed one day after infection and results are shown as the percentage of IFN- $\gamma$  $^{+}$  cells among NKT cells (mean  $\pm$  s.d.). The difference between Hexb $^{+/+}$  and Hexb $^{-/-}$  was significant for *Salmonella* ( $P = 0.001$ ). Three mice per group were analysed, and similar results obtained in two independent experiments. **c**, Bacterial burden in the lungs of J $\alpha$ 18 $^{+/+}$  and J $\alpha$ 18 $^{-/-}$  mice on days 1 (left), 3 (middle) and 5 (right) post-infection with  $5 \times 10^6$  CFU of *Spingomonas* (each bar represents mean  $\pm$  s.d. of four to five mice). Fold increase and P values are indicated. One experiment representative of three is shown. **d**, Acute lethality after inoculation of a high dose of  $5 \times 10^8$  *Spingomonas capsulata* to CD1d $^{+/+}$  versus CD1d $^{-/-}$  mice ( $n = 24$  each,  $P < 0.0001$ ). **e**, Acute serum release of IFN- $\gamma$  after infection with  $1 \times 10^7$  *Spingomonas capsulata* in CD1d $^{+/+}$  and CD1d $^{-/-}$  mutant mice. Similar results were obtained in two independent experiments. **f**, *Ehrlichia* PCR counts in spleens of CD1d $^{+/+}$  and CD1d $^{-/-}$  mice at 2 days (left) and 7 days (right) post-infection. Data represent mean  $\pm$  s.d. from three mice; one experiment representative of two is shown. Fold increase and P values are indicated.

lipids subsequently induce DC activation, probably through CD40L/CD40 interaction, as reported for  $\alpha$ GalCer<sup>15</sup>. We confirmed the early activation of NKT cells and their secretion of IFN- $\gamma$  within 24 h of infection by each of these bacteria *in vivo* (Fig. 4a). Further, we examined the response to *Salmonella* and *Sphingomonas* by NKT cells transferred into *Hexb*<sup>-/-</sup> mice lacking lysosomal iGb3, and found that the response to *Salmonella* was selectively abrogated (Fig. 4b). These results are in direct support of the proposed model of NKT cell activation by iGb3 in the case of *Salmonella* infection.

To test the central role that NKT cells might play *in vivo* against some Gram-negative, LPS-negative bacteria, we intravenously infected NKT-deficient mice and their littermate controls with *Sphingomonas*, for which the NKT cell antigens are well defined. After infection with  $1 \times 10^6$  or  $5 \times 10^6$  bacteria, both *CD1d*<sup>-/-</sup> and *J $\alpha$ 18*<sup>-/-</sup> mice had delayed bacterial clearance compared with heterozygous littermate controls, with up to 12–14-fold higher bacterial load in the lung at early time points; this demonstrates the antimicrobial function of NKT cells (Fig. 4c and Supplementary Fig. S3). Infection with a high dose of  $5 \times 10^8$  *Sphingomonas* colony-forming units (CFU) was rapidly lethal in all the wild-type mice, but in contrast, a majority of NKT-deficient mice survived (Fig. 4d and Supplementary Fig. S4). The lethal outcome in wild-type mice was associated with the explosive release of IFN- $\gamma$  and IL-12 in the serum, whereas NKT deficient mice released significantly less of the two cytokines (Fig. 4e and Supplementary Fig. S5). These results demonstrate that NKT cells mediate septic shock in response to high doses of *Sphingomonas*. Similarly, NKT-deficient mice are unable to clear *Ehrlichia* (Fig. 4f and Supplementary Fig. S6).

This study reveals two very different strategies of microbial recognition by NKT cells. For the Gram-negative, LPS-positive bacterium *Salmonella*, microbial invasion was detected by TLRs expressed on DCs, and NKT cells functioned downstream of this primary event, side by side with NK cells, through the recognition of their endogenous (self) ligand iGb3. For the Gram-negative, LPS-negative *Sphingomonas*, microbial invasion was directly detected by NKT cells through specific recognition of cell-wall glycosylceramides in the place of LPS. Although *Sphingomonas* is capable of inducing septic shock in immunocompromised humans, it is reported to be an infrequent pathogen<sup>16</sup>; however, recent findings in primary biliary cirrhosis have suggested that chronic infection might underlie severe liver inflammation associated with NKT cell redistribution<sup>17,18</sup>. Our findings suggest that similar mechanisms of direct, acute NKT cell activation by microbial cell-wall lipids may operate for other alpha-Proteobacteria, including *Ehrlichia*, a severe natural pathogen in mice and humans<sup>19</sup>. Other *Rickettsiales*, which belong to the same class of alpha-Proteobacteria, also lack peptidoglycan and LPS<sup>8,9,14</sup> and appear to be controlled by TLR2-, TLR4- and MyD88-independent mechanisms<sup>20</sup>. Thus, it is tempting to speculate that NKT cells might be specifically involved in immunity against some microbial pathogens lacking cell-wall ligands for TLRs. These novel mechanisms of NKT cell activation highlight their dual antimicrobial functions, monitoring both endogenous and exogenous glycosphingolipids in various infectious settings. □

## Methods

### Mice

*CD1d*<sup>-/-</sup> and *Trif*<sup>lps2/lps2</sup> mice were generated in our laboratories, *MyD88*<sup>-/-</sup> were from S. Akira, *J $\alpha$ 18*<sup>-/-</sup> were obtained from M. Taniguchi and *Hexb*<sup>-/-</sup> were from R. Proia. All mice were in the C57BL/6 background. In all cases, littermates obtained from heterozygous matings were genotyped by polymerase chain reaction (PCR) and used for comparative analysis. All mice were raised in a specific pathogen-free environment at the University of Chicago, according to the Institutional Animal Care and Use Committee guidelines.

### CD1d-restricted T cell responses

Lymphocyte preparations, cell staining and sorting with CD1d-glycolipid tetramers were performed as described<sup>21,22</sup>. *Griffonia simplicifolia* isolectin B4 (IB4) was from Vector Laboratories, and anti-human CD1d monoclonal antibody 51 was obtained from S. Porcelli. The human V $\alpha$ 24 NKT cell lines were derived from peripheral blood

lymphocytes (PBLs) stimulated with  $\alpha$ GalCer. Stimulation assays were performed with whole spleen cells ( $5 \times 10^5$  per 200  $\mu$ l well) or with purified T cells and antigen-presenting cells. T cells were either sorted CD1d- $\alpha$ GalCer<sup>+</sup> mouse spleen cells ( $5 \times 10^4$  per 200  $\mu$ l well), human PBLs ( $5 \times 10^5$  per 200  $\mu$ l well), obtained after Ficoll centrifugation of heparinized blood) or human NKT cell lines ( $2.5 \times 10^5$  per 200  $\mu$ l well). Antigen-presenting cells for mouse assays were bone marrow-derived DCs ( $2.5 \times 10^5$  per 200  $\mu$ l well) cultured with granulocyte-macrophage colony-stimulating factor (GM-CSF) and IL-4 (100 ng ml<sup>-1</sup> each, R&D Systems); for human assays, human irradiated allogeneic PBMCs ( $2 \times 10^5$  per 200  $\mu$ l well) were used, either fresh or cultured for 5 days with GM-CSF and IL-4 (referred to as DCs) (V $\alpha$ 24 NKT lines do not respond to allogeneic MHC antigens). Cells were washed twice and starved for 6 h in medium alone before addition to the stimulation experiments. T cells were stimulated for 48 h with the indicated lipid concentrations in 96-well round-bottom plates containing RPMI 1640 (Biofluids) supplemented with glutamine, antibiotics,  $5 \times 10^{-5}$  M 2-mercaptoethanol and 10% FCS (mouse studies) or 5% human AB serum (human studies). Concentrations of mouse and human IFN- $\gamma$  in the supernatant were measured using the respective enzyme-linked immunosorbent assay (ELISA) kits (BD Bioscience, lower detection limit of 12.5 pg ml<sup>-1</sup>).

### Bacterial strains

*Sphingomonas capsulata* (American Type Culture Collection 14666) and *Salmonella typhimurium* R71 were grown in Mueller-Hinton agar. *Ehrlichia muris* were prepared as described<sup>23</sup>. Bacteria were heat-killed by 2-h exposure to 74 °C, and  $2.5\text{--}5 \times 10^6$  CFU-equivalents per well were used for *in vitro* stimulation.

### Live infection experiments

*Sphingomonas* and *Salmonella* were grown overnight in Mueller-Hinton broth, diluted in fresh medium, grown for 8 h at 37 °C to an absorbance of 0.5 at 600 nm, washed and diluted in PBS buffer. 5–7-week-old *CD1*<sup>+/+</sup> and *CD1*<sup>-/-</sup>, or *J $\alpha$ 18*<sup>+/+</sup> and *J $\alpha$ 18*<sup>-/-</sup>, or *Hexb*<sup>+/+</sup> and *Hexb*<sup>-/-</sup> littermates were intravenously inoculated with 100  $\mu$ l of bacterial suspension. For survival experiments, dead/moribund (euthanized) mice were recorded every 2–4 h. At indicated time points after infection, bacterial counts were performed after tissue homogenization in 0.5% Triton X-100 and cultured for colony formation. For *Ehrlichia* infection experiments, mice were infected intraperitoneally with 500  $\mu$ l of a  $10^{-1}$  dilution of the *Ehrlichia muris* stock as described<sup>23</sup>.

### Determination of *Ehrlichia* cell numbers

The *Ehrlichia* load in tissues was determined by real-time PCR of the *Ehrlichia dsb* gene as described previously<sup>23</sup>. Results were normalized to GAPDH levels in the same sample and expressed as numbers of *Ehrlichia* copies per  $10^4$  GAPDH copies.

### NKT cell transfers

$5 \times 10^6$  V $\alpha$ 14 TCR transgenic thymocytes<sup>24</sup> were labelled with CFSE (carboxyfluorescein diacetate succinimidyl ester) and transferred in a volume of 50  $\mu$ l directly into the spleen of *Hexb*<sup>+/+</sup> and *Hexb*<sup>-/-</sup> littermates.

### Lipid antigens

The synthesis of *Sphingomonas* lipids (PBS 30 and PBS 59) and control PBS 50 will be described elsewhere. Synthetic compounds were characterized via <sup>1</sup>H- and <sup>13</sup>C-NMR spectroscopy and mass spectroscopy; results were consistent with the published structures of the *Sphingomonas* glycolipids. Purity was assayed using thin-layer chromatography and <sup>1</sup>H-NMR spectroscopy.

### Flow cytometry

CD1d-lipid tetramers were generated as described<sup>21</sup>. CFSE labelling and intracellular cytokine staining were performed according to manufacturer's instructions (Molecular Probes and BD Pharmingen). Anti-B220 and anti-CD69 (mouse) antibodies were from BD Pharmingen and anti-V $\alpha$ 24 (human) antibodies were from Beckman Coulter. Cells were analysed using a FACSCalibur (BD Biosciences) machine with CellQuest software.

Received 13 December 2004; accepted 24 January 2005; doi:10.1038/nature03408.

1. Park, S. H. & Bendelac, A. CD1-restricted T-cell responses and microbial infection. *Nature* **406**, 788–792 (2000).
2. Brigl, M. & Brenner, M. B. CD1: antigen presentation and T cell function. *Annu. Rev. Immunol.* **22**, 817–890 (2004).
3. Stetson, D. B. *et al.* Constitutive cytokine mRNAs mark natural killer (NK) and NK T cells poised for rapid effector function. *J. Exp. Med.* **198**, 1069–1076 (2003).
4. Nieuwenhuis, E. E. *et al.* CD1d-dependent macrophage-mediated clearance of *Pseudomonas aeruginosa* from lung. *Nature Med.* **8**, 588–593 (2002).
5. Kawakami, K. *et al.* Critical role of V $\alpha$ 14<sup>+</sup> natural killer T cells in the innate phase of host protection against *Streptococcus pneumoniae* infection. *Eur. J. Immunol.* **33**, 3322–3330 (2003).
6. Brigl, M., Bry, L., Kent, S. C., Gumperz, J. E. & Brenner, M. B. Mechanism of CD1d-restricted natural killer T cell activation during microbial infection. *Nature Immunol.* **4**, 1230–1237 (2003).
7. Fischer, K. *et al.* Mycobacterial phosphatidylinositol mannoside is a natural antigen for CD1d-restricted T cells. *Proc. Natl Acad. Sci. USA* **101**, 10685–10690 (2004).
8. Lin, M. & Rikhiya, Y. *Ehrlichia chaffeensis* and *Anaplasma phagocytophilum* lack genes for lipid A biosynthesis and incorporate cholesterol for their survival. *Infect. Immun.* **71**, 5324–5331 (2003).
9. Kawahara, K., Moll, H., Knirel, Y. A., Seydel, U. & Zahringer, U. Structural analysis of two glycosphingolipids from the lipopolysaccharide-lacking bacterium *Sphingomonas capsulata*. *Eur. J. Biochem.* **267**, 1837–1846 (2000).
10. Hoebe, K. *et al.* Identification of Lps2 as a key transducer of MyD88-independent TIR signalling. *Nature* **424**, 743–748 (2003).
11. Zhou, D. *et al.* Lysosomal glycosphingolipid recognition by NKT cells. *Science* **306**, 1786–1789 (2004).



12. Proia, R. L. Glycosphingolipid functions: insights from engineered mouse models. *Phil. Trans. R. Soc. Lond. B* **358**, 879–883 (2003).
13. Keusch, J. J., Manzella, S. M., Nyame, K. A., Cummings, R. D. & Baenziger, J. U. Expression cloning of a new member of the ABO blood group glycosyltransferases, iGb3 synthase, that directs the synthesis of isoglobos-glycosphingolipids. *J. Biol. Chem.* **275**, 25308–25314 (2000).
14. Amano, K. *et al.* Deficiency of peptidoglycan and lipopolysaccharide components in *Rickettsia tsutsugamushi*. *Infect. Immun.* **55**, 2290–2292 (1987).
15. Fujii, S., Liu, K., Smith, C., Bonito, A. J. & Steinman, R. M. The linkage of innate to adaptive immunity via maturing dendritic cells *in vivo* requires CD40 ligation in addition to antigen presentation and CD80/86 costimulation. *J. Exp. Med.* **199**, 1607–1618 (2004).
16. Hsueh, P. R. *et al.* Nosocomial infections caused by *Sphingomonas paucimobilis*: clinical features and microbiological characteristics. *Clin. Infect. Dis.* **26**, 676–681 (1998).
17. Kita, H. *et al.* Quantitation and phenotypic analysis of natural killer T cells in primary biliary cirrhosis using a human CD1d tetramer. *Gastroenterology* **123**, 1031–1043 (2002).
18. Selmi, C. *et al.* Patients with primary biliary cirrhosis react against a ubiquitous xenobiotic-metabolizing bacterium. *Hepatology* **38**, 1250–1257 (2003).
19. Olano, J. P. & Walker, D. H. Human ehrlichiosis. *Med. Clin. North Am.* **86**, 375–392 (2002).
20. von Loewenich, F. D., Scorpio, D. G., Reischl, U., Dumler, J. S. & Bogdan, C. Frontline: control of *Anaplasma phagocytophilum*, an obligate intracellular pathogen, in the absence of inducible nitric oxide synthase, phagocyte NADPH oxidase, tumor necrosis factor, Toll-like receptor (TLR)2 and TLR4, or the TLR adaptor molecule MyD88. *Eur. J. Immunol.* **34**, 1789–1797 (2004).
21. Benlagha, K., Weiss, A., Beavis, A., Teyton, L. & Bendelac, A. *In vivo* identification of glycolipid antigen specific T cells using fluorescent CD1d tetramers. *J. Exp. Med.* **191**, 1895–1903 (2000).
22. Lee, P. T., Benlagha, K., Teyton, L. & Bendelac, A. Distinct functional lineages of human V $\alpha$ 24 natural killer T cells. *J. Exp. Med.* **195**, 637–641 (2002).
23. Ismail, N. *et al.* Overproduction of TNF- $\alpha$  by CD8 $^{+}$  type 1 cells and down-regulation of IFN- $\gamma$  production by CD4 $^{+}$  Th1 cells contribute to toxic shock-like syndrome in an animal model of fatal monocytotropic ehrlichiosis. *J. Immunol.* **172**, 1786–1800 (2004).
24. Bendelac, A., Hunziker, R. D. & Lantz, O. Increased interleukin 4 and immunoglobulin E production in transgenic mice overexpressing NK1 T cells. *J. Exp. Med.* **184**, 1285–1293 (1996).

Supplementary Information accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. Wei for reading the manuscript, K. Thompson for help with biochemical characterization of *Sphingomonas* and growth of bacteria, S. Porcelli for the gift of anti-human CD1d, R. Duggan, J. Marvin and B. Eisfelder for cell sorting, L. Taylor for managing the mouse colonies, and the University of Chicago Digestive Disease Research center for equipment. Supported by NIH grants to A.B., P.S.B. and L.T., an NIH award to K.L.D. and O.S., an NIH grant to B.B., a Cancer Research Institute fellowship to J.M. and D.Z., and a fellowship from the Fondation pour la Recherche Médicale to P.S.-M.

**Authors' contributions** P.B.S. and A.B. are co-senior authors.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence and requests for materials** should be addressed to A.B. (abendela@bsd.uchicago.edu) or P.B.S. (paul\_savage@byu.edu).

## Fission yeast Mes1p ensures the onset of meiosis II by blocking degradation of cyclin Cdc13p

Daisuke Izawa<sup>1</sup>, Masuo Goto<sup>3\*</sup>, Akira Yamashita<sup>2</sup>, Hiroyuki Yamano<sup>4</sup> & Masayuki Yamamoto<sup>1,2</sup>

<sup>1</sup>Department of Biophysics and Biochemistry, Graduate School of Science, and  
<sup>2</sup>Molecular Genetics Research Laboratory, University of Tokyo, Hongo, Tokyo 113-0033, Japan

<sup>3</sup>Division of Cell Proliferation, National Institute for Basic Biology, Okazaki, Aichi 444-0867, Japan

<sup>4</sup>Cell Cycle Control Group, Marie Curie Research Institute, The Chart, Oxted, Surrey RH8 0TL, UK

\* Present address: NIEHS, NIH, Research Triangle Park, North Carolina 27709-2233, USA

Meiosis is a special form of nuclear division to generate eggs, sperm and spores in eukaryotes. Meiosis consists of the first (MI) and the second (MII) meiotic divisions, which occur consecutively. MI is reductional, in which homologous chromosomes derived from parents segregate. MI is supported by an elaborate mechanism involving meiosis-specific cohesin and its protector<sup>1</sup>.

MII is equational, in which replicated sister-chromatids separate as in mitosis. MII is generally considered to mimic mitosis in mechanism. However, fission yeast Mes1p is essential for MII but dispensable for mitosis. The *mes1-B44* mutant arrests before MII<sup>2</sup>. Transcription of *mes1* is low in vegetative cells and boosted in a narrow window between late MI and late MII<sup>3</sup>. The *mes1* mRNA undergoes meiosis-specific splicing<sup>4</sup>. Here we show that Mes1p is a factor that suppresses the degradation of cyclin Cdc13p at anaphase I. Mes1p binds to Slp1p, an activator of APC/C (anaphase promoting complex/cyclosome), and counteracts its function to engage Cdc13p in proteolysis. Inhibition of APC/C-dependent degradation of Cdc13p by Mes1p was reproduced in a *Xenopus* egg extract. We therefore propose that Mes1p has a key function in saving a sufficient level of MPF (M-phase-promoting factor) activity required for the execution of MII.

MPF is a complex of Cdc2 kinase and M-phase cyclin Cdc13p. To investigate MPF dynamics during meiosis in wild-type and *mes1Δ* cells, we analysed zygotes (*h<sup>90</sup>/h<sup>90</sup>*) expressing Cdc13p tagged with green fluorescent protein (GFP) at various stages of meiosis (Fig. 1a–c). Figure 1a shows a profile of Cdc13p-GFP in wild-type zygotes. Each cell was assigned to a specific meiotic phase according to the number of nuclei and the length of the spindle(s) revealed by CFP-tubulin. Typical examples for metaphase I, anaphase I, metaphase II and anaphase II are displayed. At metaphase I, Cdc13p-GFP was detected in the nucleus, especially on the spindle. It remained in the nucleus until anaphase I. This contrasted with mitosis, during which Cdc13p disappears at anaphase<sup>5</sup>. At metaphase II, Cdc13p-GFP was again detected on the spindle. It disappeared completely from the nucleus at anaphase II. These observations confirmed a previous report that MPF activity is maintained from the initiation of MI to the end of MII in fission yeast<sup>6</sup>. *mes1Δ* zygotes showed the same Cdc13p-GFP localization until metaphase I, but all Cdc13p-GFP disappeared from the nucleus at anaphase I (Fig. 1b). Quantitative analysis of Cdc13p-GFP fluorescence per cell indicated that wild-type cells kept half of Cdc13p remaining at anaphase I, whereas *mes1Δ* cells lost most of it (Fig. 1c).

We then measured Cdc2 kinase activity during synchronous meiosis induced by thermal inactivation of Pat1 kinase in pre-starved homozygous (*h<sup>-</sup>/h<sup>-</sup>*) diploid cells, using histone H1 as a substrate. We monitored the progression of meiosis by counting the number of nuclei in a cell (Fig. 1d, e). In wild-type cells, the activity increased at the onset of MI (3.75 h) and remained high until early MII (4.5 h) (Fig. 1d, f), as reported<sup>6</sup>. In contrast, the activity seemed to decrease after MI in *mes1Δ* cells (Fig. 1e, f). In immunoblotting, Cdc13p disappeared about 20 min earlier in *mes1Δ* than in wild-type cells (Fig. 1d, e). Mes1p tagged with three copies of a haemagglutinin epitope (3HA), expressed from the authentic *mes1* promoter, reached its peak in about late MI/early MII (Fig. 1d). Cdc2p was equally abundant throughout meiosis in *mes1Δ* and wild-type cells, and its phosphorylation at tyrosine 15, a measure of the kinase activity<sup>7</sup>, was indistinguishable between them (Fig. 1d,e).

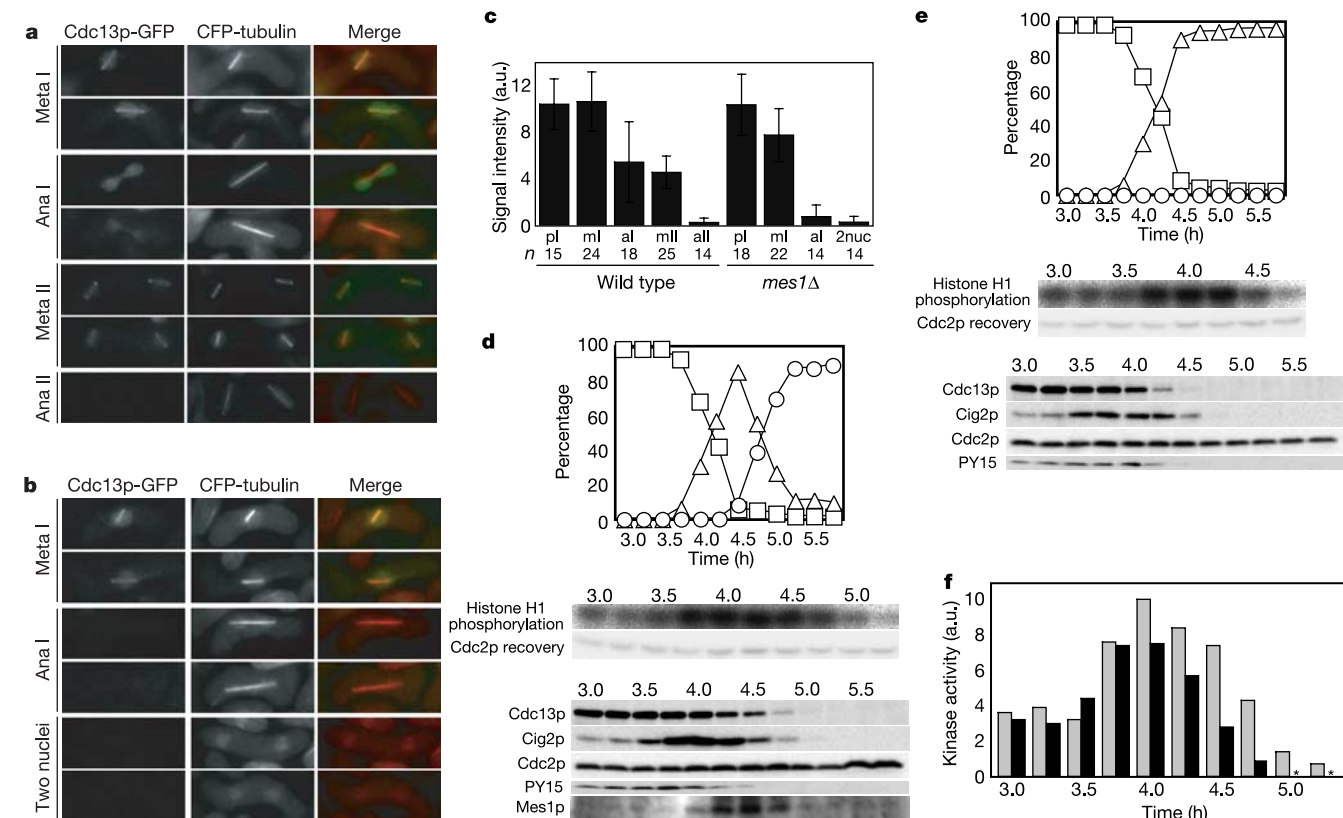
We attempted to raise the level of Cdc13p in *mes1Δ* cells by using a plasmid pPmes-*cdc13*, which carried the *cdc13* open reading frame under the control of the *mes1* promoter. *mes1Δ* cells transformed with pPmes-*cdc13* produced four-spored asci efficiently, whereas control cells carrying the vector pPmes were arrested at MII (Fig. 2a). We found that pPmes-*cig2*, carrying the *cig2* open reading frame<sup>8</sup>, could also suppress *mes1Δ* (Fig. 2a). The *cig2* gene, encoding a G1/S cyclin, is known to be expressed at about MII, although it is not strictly essential for MII<sup>6</sup>. Immunoblot analysis indicated that Cig2p was slightly less abundant and disappeared earlier in *mes1Δ* cells (Fig. 1d, e). Efficient suppression of *mes1Δ* by the supply of either Cdc13p or Cig2p suggested that the central role of Mes1p might be to maintain cyclin between MI and MII.

In a genetic screen we isolated a truncated form of *slp1* as a multi-copy suppressor of *mes1-B44*. Slp1p is a WD-repeat protein activating APC/C, which is homologous to budding yeast Cdc20p<sup>9–12</sup>. Cdc20p recruits M-phase cyclin and securin to APC/C<sup>13</sup>. The suppressor clone lacked the carboxy-terminal 35 residues. We examined a series of truncated Slp1p for the ability to suppress *mes1Δ* (Fig. 2b). Slp1p carrying either residues 1–334 or residues 1–384 or residues 1–420 or residues 1–453 could suppress *mes1Δ*. Interestingly, the full-length form (1–488) failed to suppress *mes1Δ*, as did the shortest form examined (1–250).

The original *slp1* mutant allele, *slp1-362*, carries a nonsense mutation at codon 376 and causes temperature-sensitive growth, whereas disruption of *slp1* is lethal<sup>9,14</sup>. Both *slp1Δ* and *slp1-362* cells show non-segregated chromosomes and metaphase arrest in mitosis as a terminal phenotype, possibly due to a failure in degrading securin Cut2p<sup>10,15</sup>. None of the above Slp1p deletions complemented *slp1-362*, whereas the full-length form did. Moreover, we noticed that deletion clones capable of *mes1Δ* suppression inhibited growth of *slp1-362* cells at the non-restrictive temperature (data not shown). This suggested that the Slp1p deletions were non-functional and could interfere with functional Slp1p. As Slp1p(1–375) generated from the *slp1-362* allele was unlikely to be functional, this nonsense allele was presumed to give a small quantity of

read-through products, which could support growth of the mutant cells at low temperature. The *slp1-362* mutant was defective in meiosis even at low temperature: zygotes were arrested largely at MII and eventually generated one-spored or two-spored asci (Fig. 2c). However, the *slp1-362 mes1Δ* double mutant was fairly competent in meiosis, generating more four-spored asci than either single mutant (Fig. 2c). These observations indicated that reduction of Slp1p activity and loss of *mes1* function might compensate for each other.

Tagged Mes1p and tagged Slp1p could be co-immunoprecipitated from cell extracts expressing them (Fig. 2d). A pull-down assay using Mes1p tagged with glutathione S-transferase (GST) and truncated Slp1p-3HA indicated that Slp1p(1–453), which could suppress *mes1Δ*, did not bind to Mes1p (Fig. 2e). Slp1p(166–488), composed of the entire WD repeats and the C-terminal 15 residues, retained weak affinity for Mes1p, but other deletions examined did not (Fig. 2e). Thus, the complete WD repeats and the C-terminal residues seemed to be necessary for the binding of Slp1p to Mes1p. Binding of truncated Slp1p-3HA to GST-Cdc13p gave essentially the same pattern (Fig. 2e), indicating that Mes1p and Cdc13p might compete for the same binding site on Slp1p. Deletion analysis indicated that the amino-terminal half of Mes1p, corresponding to the first exon (58 amino acids), could bind to Slp1p and perform



**Figure 1** Mes1p is required to maintain MPF activity after MI. **a**, Dynamics of Cdc13p in wild-type cells undergoing zygotic meiosis, revealed by GFP tagging. Cells were classified by the appearance of their spindles revealed by CFP-Atb2p ( $\alpha$ -tubulin). In merged panels, green represents Cdc13p and red represents tubulin. Ana, anaphase; meta, metaphase. **b**, Dynamics of Cdc13p in *mes1Δ* cells, analysed similarly to **a**. **c**, Intensity of Cdc13p-GFP fluorescence emitted from a cell at various phases. p, prophase; m, metaphase; a, anaphase; n, the number of cells analysed. Error bars show standard deviation. **d**, **e**, The MPF activity and the abundance of MPF components were measured during *pat1*-induced synchronous meiosis in wild-type (**d**) and *mes1Δ* (**e**) cells.

Samples were taken every 15 min from the onset of MI. To monitor the progression of meiosis, each sample was examined microscopically for the number of nuclei in a cell (top) (squares, one nucleus; triangles, two nuclei; circles, four nuclei). Cdc2p was pulled down and assayed for histone H1 kinase activity (middle). Amounts of Cdc13p, Cig2p and Cdc2p, and phosphorylation of Cdc2p at tyrosine 15 were determined by western blotting (bottom). Mes1p in wild-type cells was also detected (**d**). **f**, Histone H1 kinase activities detected in **d** (grey) and **e** (black) are quantified and shown graphically. Asterisk, not examined; a.u., arbitrary units.

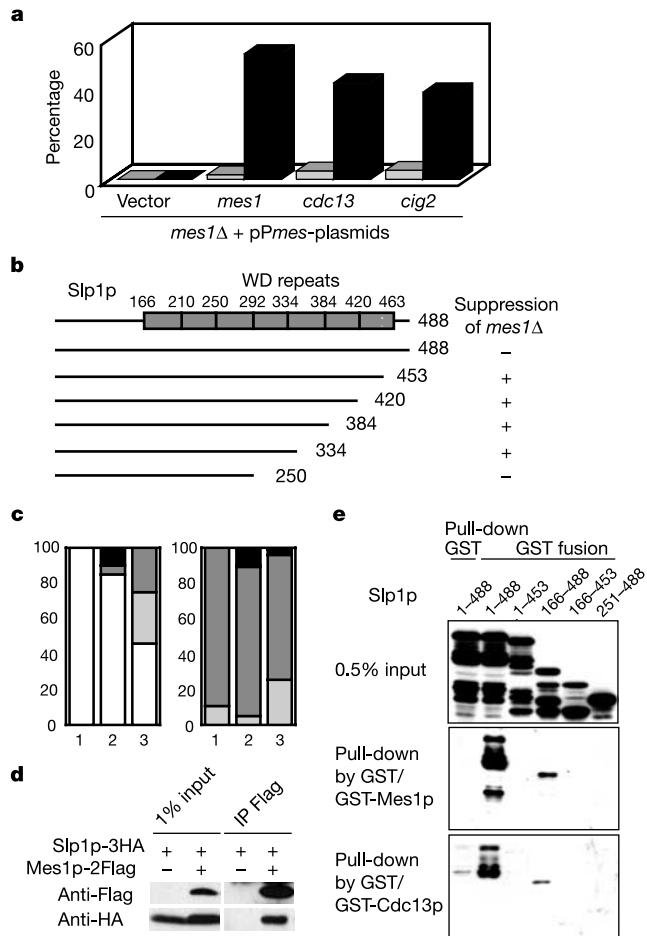
its essential function (data not shown). This region contained a putative KEN box and a possible D-box.

We confirmed counteraction between Mes1p and Slp1p in mitotic cells in three ways. First, overproduction of Mes1p was lethal, and so was overproduction of Slp1p, but cells overproducing both proteins could survive (data not shown). Second, overproduction of Mes1p increased nearly tenfold the portion of cells that exhibited GFP-tagged Cut2p on the mitotic spindle (Fig. 3a), but overproduction of Slp1p decreased Cut2p in the nucleus (data not shown). Last, as shown below, Mes1p overproduction could cause a

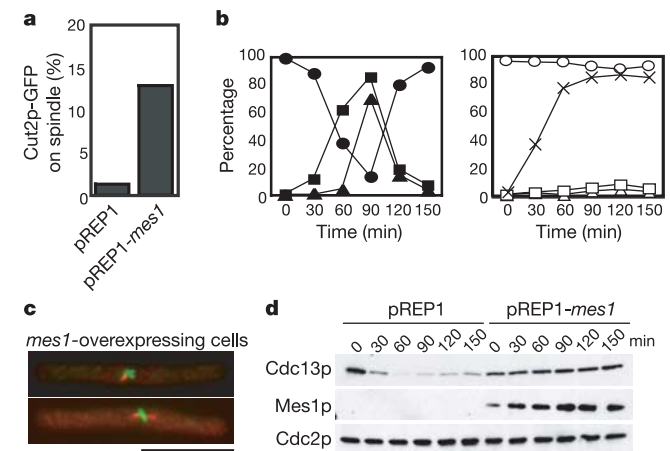
failure in the metaphase–anaphase transition in mitosis, like either the *slp1-362* mutation<sup>9</sup> or the overproduction of Mad2p, a spindle checkpoint protein that blocks the function of Slp1p<sup>16</sup>. *cdc25-22* cells arrest in G2 at the restrictive temperature and enter M phase synchronously after shift-down<sup>17</sup> (Fig. 3b, left). We overproduced Mes1p in G2-arrested *cdc25-22* cells. After shift-down, these cells accumulated condensed chromosomes and failed to segregate them (Fig. 3b, right). They carried either no detectable spindle (two-thirds) or short spindles indicative of metaphase arrest (one-third) (Fig. 3c). A high level of Cdc13p was maintained in them, whereas control cells exhibited obvious degradation of Cdc13p at about the onset of nuclear division (Fig. 3d). Whether Cut2p was also stabilized in Mes1p-overproducing cells was inconclusive, because quantities of Cut2p were not remarkably decreased even in the control cells performing mitosis (estimated to be 50% in the original report<sup>15</sup>).

All the above observations strongly support that Mes1p is an inhibitor of Slp1p. To investigate the inhibitory mechanism, we examined interaction of Slp1p and its substrate Cdc13p in the presence or absence of Mes1p. We added bacterially purified GST–Cdc13p to a fission yeast cell extract containing Slp1p–3HA, with either maltose-binding protein (MBP) or MBP–Mes1p. When MBP was added to the extract, Slp1p–3HA could be pulled down by GST–Cdc13p, but when MBP–Mes1p was added in a 3–15-fold molar excess over GST–Cdc13p, the interaction was disrupted in a dose-dependent manner (Fig. 4a). We obtained essentially the same results when GST–Cut2p replaced GST–Cdc13p (data not shown). We therefore speculate that Mes1p inhibits the function of Slp1p by sequestering it from the substrates, probably in a competitive manner.

Apart from Slp1p, two APC/C activators involved in the degradation of Cdc13p have been reported in fission yeast. They are Ste9p/Srw1p, implicated in Cdc13p degradation in G1, and Mfr1p, active at anaphase II<sup>18</sup>. Because the former has been inferred to have

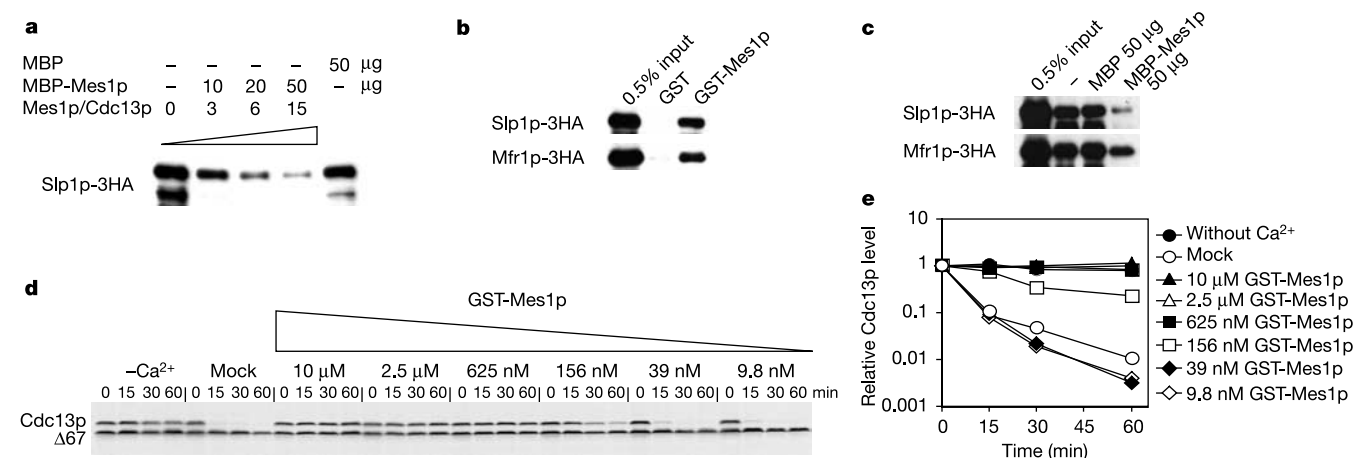


**Figure 2** Interaction of Mes1p with Slp1p, which recruits Cdc13p to proteolysis. **a**, Suppression of *mes1Δ* by the overproduction of cyclin at MII. Homothallic haploid cells transformed with pPmes (vector), pPmes-*mes1* (positive control), pPmes-*cdc13* or pPmes-*cig2*, were incubated on sporulation medium SPA at 30 °C for 18 h, and spores generated in zygotes were scored: grey columns, one or two spores; black columns, three or four spores. **b**, Schematic diagram of suppression of *mes1Δ* by the overproduction of truncated Slp1p. **c**, Suppression of *mes1Δ* by the *slp1-362* mutation. Each single mutant and the double mutant were incubated on SPA at 26.5 °C to induce mating and meiosis. The number of spores and that of nuclei in zygotes were scored after 24 h. Lane 1, *mes1Δ*; lane 2, *mes1Δ slp1-362*; lane 3, *slp1-362*. In the left panel the number of spores is indicated by white (none), light grey (one), dark grey (two) or black (three or four). In the right panel the number of nuclei are indicated by light grey (one), dark grey (two) or black (three or four). **d**, Physical interaction of Slp1p with Mes1p *in vivo*. Cell extracts containing Slp1p-3HA and either Mes1p-2Flag or 2Flag were subjected to immunoprecipitation with anti-Flag and analysed by western blotting. **e**, Interaction of truncated Slp1p with Mes1p or Cdc13p. Extracts prepared from wild-type cells producing each 3HA-tagged truncated Slp1p, as indicated in the panel, were mixed with GST-Mes1p, GST-Cdc13p or control GST, and subjected to pull-down assay followed by western blotting.



**Figure 3** Mes1p inhibits the function of Slp1p in activating APC/C. **a**, The proportion exhibiting Cut2p-GFP on mitotic spindles was compared between *mes1*-overexpressing (pREP1-*mes1*) and control (pREP1) cells. **b**, The number of nuclei, the septum index and the presence of condensed chromosomes were chased in synchronized *cdc25-22* cells, either carrying the vector (pREP1; left) or overexpressing *mes1* (right). Sampling was performed every 30 min. Ellipses, one nucleus; squares, two nuclei; triangles, septum; crosses, condensed chromosomes. **c**, Typical *mes1*-overexpressing cells, showing a short spindle. Cells in **b** at 60 min were subjected to immunostaining with anti-tubulin antibody TAT-1 (red) and DNA-staining with 4,6-diamidino-2-phenylindole (green). Scale bar, 10 μm. **d**, The same series of samples as in **b** were disrupted with glass beads, and proteins were separated by SDS-PAGE. Cdc13p, Mes1p and Cdc2p were detected by western blotting.





**Figure 4** Mes1p blocks Slp1p-APC/C function to degrade Cdc13p. **a**, Purified GST-Cdc13p was mixed with a cell extract containing Slp1p-3HA, together with the indicated amount of either MBP-Mes1p or MBP. GST-Cdc13p was then pulled down and co-precipitation of Slp1p-3HA was analysed by western blotting. **b**, Binding of Mes1p to Mfr1p was examined by pull-down assay as in Fig. 2e. **c**, Inhibition of Mfr1p-Cdc13p interaction by Mes1p was examined as in **a**. **d**, Influence of Mes1p on the destruction of

Cdc13p in *Xenopus* egg extracts. Cdc13p and Cdc13p( $\Delta$ 67) were translated together to provide destructible and indestructible substrates. GST-Mes1p was added at the indicated concentrations. Mock, only phosphate-buffered saline was added. CaCl<sub>2</sub> was added to all reactions except the leftmost four lanes. **e**, Graphs quantifying the band intensities shown in **d**.

little influence on meiosis<sup>18</sup>, we investigated the possible interaction of Mes1p with Mfr1p. In a pull-down assay Mes1p bound to Mfr1p (Fig. 4b), and in a competition assay Mes1p seemed to inhibit the interaction between Mfr1p and Cdc13p, although the inhibition might be less effective than against Slp1p (Fig. 4c).

Finally, to examine whether Mes1p could inhibit APC/C activity *in vitro*, we performed an APC/C-dependent Cdc13p destruction assay with *Xenopus* egg extracts, as described previously<sup>19</sup>. The results obtained were remarkably clear (Fig. 4d, e). When phosphate-buffered saline (mock; leftmost eight lanes in panel e) or GST (not shown) was added, Cdc13p was destroyed in a Ca<sup>2+</sup>-dependent manner. However, when GST-Mes1p was added, proteolysis of Cdc13p was suppressed in proportion to the dose of GST-Mes1p. These results provide unambiguous biochemical evidence that Mes1p is an APC/C inhibitor.

Thus, we propose the following model for meiosis II. Fission yeast produces Mes1p to inhibit the complete destruction of cyclin by APC/C-proteasome at anaphase I. Spared Cdc13p, with Cig2p expressed under meiosis-specific regulation<sup>6</sup>, secures the MPF activity necessary for the initiation of MII. Mes1p persists only transiently, and when Mes1p is removed by a mechanism not yet explained, Slp1p recruits Cdc13p effectively to proteolysis and promotes anaphase II. Cig2p is a substrate of APC/C<sup>20</sup>, and Slp1p might also mediate its destruction here, although this is not critically proven. The degradation of cyclin in MII (deduced from Fig. 1d and ref. 18), which is probably backed up by newly synthesized Mfr1p<sup>18</sup>, seems to occur earlier relative to the execution of nuclear division than that in MI (deduced from Fig. 1e). This earlier degradation might be necessary to facilitate sporulation, because loss of *mfr1* causes a delay in spore formation<sup>18</sup>.

There have been reports of maintenance of residual MPF activity after MI in *Xenopus*<sup>21,22</sup>. However, in *Xenopus* this control seems to rely on downregulation of the activity of Wee1/Myt1 kinase, which phosphorylates Cdc2 kinase on tyrosine 15. Furthermore, an artificial decrease in the MPF activity after MI brought on by an increase in Wee1 can provoke DNA replication in the frog<sup>21,22</sup>. It is unclear why we do not see DNA replication in fission yeast in a similar situation brought on by the absence of Mes1p. One possibility is that the ectopic DNA replication between MI and MII might require the phosphorylation of some critical factor by Wee1 kinase.

Another possibility is that the meiotic cell cycle of *Xenopus* eggs might be more plastic than that in yeast, because oogenesis should furnish a pause for fertilization after the completion of MI, whereas yeast has no such need.

Mes1p seems to be a novel type of APC/C inhibitor. Two APC/C inhibitors, namely Mad2p and Emi1, have been described. Mad2p is responsible for spindle checkpoint control and is widely conserved among eukaryotes<sup>23,24</sup>. Emi1, identified so far in animals, functions as an early mitotic inhibitor<sup>25</sup> and might<sup>26</sup> or might not<sup>27</sup> regulate MII arrest. These two proteins directly inhibit Cdc20p/Slp1p but in different fashions<sup>28</sup>. Mes1p seems distinct from Mad2p, because they share no obvious homology, function in different processes and apparently bind to different sites on Slp1p (ref. 23 and this study). Indeed, the artificial expression of Mad2p after MI did not suppress *mes1* $\Delta$  (data not shown). However, the relationship between Mes1p and Emi1 might deserve closer investigation, especially because the meiotic function of Emi1 is currently controversial<sup>26,27</sup>. □

## Methods

### General methods and plasmids

Methods of handling *Schizosaccharomyces pombe* were described previously<sup>29</sup>. Plasmid pPmes was constructed by replacing the *nmf1* promoter of pREP1 (ref. 30) with a 1-kilobase DNA fragment carrying the *mes1* promoter.

### Observation of GFP-tagged Cdc13p and Cut2p

Homothallic haploid cells expressing Cdc13p-GFP from the *cdc13* promoter were transformed with pREP81-CFP-*atb2* to reveal microtubules and grown to exponential phase in MM + N medium at 25 °C. They were washed, spotted on sporulation medium SPA and incubated at 25 °C for 8 h to induce mating and meiosis. Live images of zygotes were monitored at 24–26 °C under a fluorescence microscope (Axioplan2; Zeiss) equipped with a cooled charge-coupled-device camera (Quantix; Photometrics) and Metamorph software (Universal Imaging Corporation).

To observe effects of Mes1p on Cut2p, homothallic haploid cells expressing *cut2-GFP* were transformed with either pREP1-*mes1*, which expresses *mes1* in the absence of thiamine<sup>30</sup>, or vector pREP1. The transformants were cultured in thiamine-rich medium SD at 30 °C to early exponential phase. They were transferred to thiamine-free MM medium, and GFP signals visible on mitotic spindles were scored after 15 h, when full induction of *mes1* expression was expected, with the same optical equipment as above.

### Histone H1 kinase assay

Synchronous meiosis driven by *pat1-114* was induced in pre-starved diploid cells. Cells were disrupted with glass beads in HB buffer. An extract containing 0.5 mg of protein was incubated at 4 °C for 30 min with 20  $\mu$ l of p13<sup>suc1</sup> agarose beads (Upstate). After being washed three times with HB buffer, the beads were incubated with 20  $\mu$ l of kinase reaction

mix (1 mg ml<sup>-1</sup> histone H1 (Calbiochem), 2 µCi of [γ-<sup>32</sup>P]ATP (Amersham) and 50 µM unlabelled ATP) at 30 °C for 10 min. The reaction was terminated by being boiled for 3 min with 20 µl of 2 × sample buffer, and samples were analysed by SDS–PAGE. Phosphorylation of histone H1 was detected by autoradiography and quantified with an ImageAnalyzer FLA2000 (Fuji Film).

### Western blotting

Sample preparation and subsequent analysis were performed as described<sup>6</sup>. We used the following antibodies: anti-Cdc13p 6F at a dilution of 1:1,000 (a gift from P. Nurse), anti-Cig2p 3A11 at a dilution of 1:1,000 (a stock of H.Y.), anti-Mes1p at a dilution of 1:1,000 (this study), anti-Cdk1/Cdc2 PSTAIR at a dilution of 1:2,000 (Upstate), anti-PY15 at a dilution of 1:1,000 (New England Biolabs), monoclonal anti-HA 12CA5 at a dilution of 1:3,000 (Roche), polyclonal anti-HA Y-11 at a dilution of 1:3,000 (Santa Cruz Biotechnology), and anti-Flag M2 at a dilution of 1:3,000 (Sigma).

### Immunoprecipitation and pull-down assay

To see the physical interaction between Slp1p and Mes1p *in vivo*, a strain carrying *slp1-3HA* was transformed with either vector *pFlag1* or *pFlag1-mes1*. Extracts prepared from the transformants were subjected to immunoprecipitation with anti-Flag. For the pull-down assay, GST-fused and MBP-fused proteins were produced bacterially and purified in accordance with the standard protocols. GST-fused Cdc13p trapped on glutathione beads was mixed with extracts prepared from cells expressing Slp1p-3HA, together with either MBP or MBP-Mes1p. Immunoprecipitated or pulled-down proteins were analysed by western blotting.

### Synchronization of mitosis using the *cdc25-22* mutation

A *cdc25-22* strain<sup>17</sup> was transformed with either *pREP1-mes1* or *pREP1*, and the transformants were cultured in SD medium at 25 °C to early exponential phase. They were transferred to MM medium and incubated for 15 h to derepress *mes1* expression; they were then kept at 36 °C for 3.5 h and shifted down to 25 °C to restart the cell cycle from G2. Samples were taken every 30 min from the shift-down until completion of the subsequent mitosis in the control vector-carrying cells, and subjected to analyses.

### *Xenopus* egg extracts and cyclin destruction assay

Preparation of the extracts and assay procedures were described previously<sup>19</sup>. Fission yeast Cdc13p and Cdc13p(Δ67), an indestructible variant lacking the N-terminal 67 residues, were labelled with <sup>35</sup>S and used as substrates. Bacterially expressed and purified GST-Mes1p or GST was added to some assays.

Received 17 November 2004; accepted 27 January 2005; doi:10.1038/nature03406.

1. Watanabe, Y. Modifying sister chromatid cohesion for meiosis. *J. Cell Sci.* **117**, 4017–4023 (2004).
2. Shimoda, C., Hirata, A., Kishida, M., Hashida, T. & Tanaka, K. Characterization of meiosis-deficient mutants by electron microscopy and mapping of four essential genes in the fission yeast *Schizosaccharomyces pombe*. *Mol. Gen. Genet.* **200**, 252–257 (1985).
3. Mata, J., Lyne, R., Burns, G. & Bahler, J. The transcriptional program of meiosis and sporulation in fission yeast. *Nature Genet.* **32**, 143–147 (2002).
4. Kishida, M., Nagai, T., Nakaseko, Y. & Shimoda, C. Meiosis-dependent mRNA splicing of the fission yeast *Schizosaccharomyces pombe mes1<sup>+</sup>* gene. *Curr. Genet.* **25**, 497–503 (1994).
5. Decottignies, A., Zarzov, P. & Nurse, P. *In vivo* localisation of fission yeast cyclin-dependent kinase cdc2p and cyclin B cdc13p during mitosis and meiosis. *J. Cell Sci.* **114**, 2627–2640 (2001).
6. Borgne, A., Murakami, H., Aye, J. & Nurse, P. The G1/S cyclin Cig2p during meiosis in fission yeast. *Mol. Biol. Cell* **13**, 2080–2090 (2002).
7. Gould, K. L. & Nurse, P. Tyrosine phosphorylation of the fission yeast *cdc2<sup>+</sup>* protein kinase regulates entry into mitosis. *Nature* **342**, 39–45 (1989).
8. Bueno, A. & Russell, P. Two fission yeast B-type cyclins, *cig2* and *cdc13*, have different functions in mitosis. *Mol. Cell Biol.* **13**, 2286–2297 (1993).
9. Matsumoto, T. A fission yeast homolog of CDC20/p55CDC/Fizzy is required for recovery from DNA damage and genetically interacts with p34<sup>cdc2</sup>. *Mol. Cell Biol.* **17**, 742–750 (1997).
10. Visintin, R., Prinz, S. & Amon, A. *CDC20* and *CDH1*: a family of substrate-specific activators of APC-dependent proteolysis. *Science* **278**, 460–463 (1997).
11. Fang, G., Yu, H. & Kirschner, M. W. Direct binding of CDC20 protein family members activates the anaphase-promoting complex in mitosis and G1. *Mol. Cell* **2**, 163–171 (1998).
12. Peters, J. M. The anaphase-promoting complex: proteolysis in mitosis and beyond. *Mol. Cell* **9**, 931–943 (2002).
13. Burton, J. L. & Solomon, M. J. D box and KEN box motifs in budding yeast Hsl1p are required for APC-mediated degradation and direct binding to Cdc20p and Cdh1p. *Genes Dev.* **15**, 2381–2395 (2001).
14. Yamada, H. Y., Matsumoto, S. & Matsumoto, T. High dosage expression of a zinc finger protein, Grl1, suppresses a mutant of fission yeast *slp1<sup>+</sup>*, a homolog of CDC20/p55CDC/Fizzy. *J. Cell Sci.* **113**, 3989–3999 (2000).
15. Funabiki, H. *et al.* Cut2 proteolysis required for sister-chromatid separation in fission yeast. *Nature* **381**, 438–441 (1996).
16. He, X., Patterson, T. E. & Sazer, S. The *Schizosaccharomyces pombe* spindle checkpoint protein mad2p blocks anaphase and genetically interacts with the anaphase-promoting complex. *Proc. Natl Acad. Sci. USA* **94**, 7965–7970 (1997).
17. Moreno, S., Hayles, J. & Nurse, P. Regulation of p34<sup>cdc2</sup> protein kinase during mitosis. *Cell* **58**, 361–372 (1989).
18. Blanco, M. A., Pelloquin, L. & Moreno, S. Fission yeast mfr1 activates APC and coordinates meiotic nuclear division with sporulation. *J. Cell Sci.* **114**, 2135–2143 (2001).
19. Yamano, H., Gannon, J., Mahbubani, H. & Hunt, T. Cell cycle-regulated recognition of the destruction box of cyclin B by the APC/C in *Xenopus* egg extracts. *Mol. Cell* **13**, 137–147 (2004).
20. Yamano, H. *et al.* The spike of S phase cyclin Cig2 expression at the G1-S border in fission yeast requires both APC and SCF ubiquitin ligases. *Mol. Cell* **6**, 1377–1387 (2000).

21. Nakajo, N. *et al.* Absence of Wee1 ensures the meiotic cell cycle in *Xenopus* oocytes. *Genes Dev.* **14**, 328–338 (2000).
22. Iwabuchi, M., Ohsumi, K., Yamamoto, T. M., Sawada, W. & Kishimoto, T. Residual Cdc2 activity remaining at meiosis I exit is essential for meiotic M-M transition in *Xenopus* oocyte extracts. *EMBO J.* **19**, 4513–4523 (2000).
23. Kim, S. H., Lin, D. P., Matsumoto, S., Kitazono, A. & Matsumoto, T. Fission yeast Slp1: an effector of the Mad2-dependent spindle checkpoint. *Science* **279**, 1045–1047 (1998).
24. Shah, J. V. & Cleveland, D. W. Waiting for anaphase: Mad2 and the spindle assembly checkpoint. *Cell* **103**, 997–1000 (2000).
25. Reimann, J. D. *et al.* Emi1 is a mitotic regulator that interacts with Cdc20 and inhibits the anaphase promoting complex. *Cell* **105**, 645–655 (2001).
26. Reimann, J. D. & Jackson, P. K. Emi1 is required for cytostatic factor arrest in vertebrate eggs. *Nature* **416**, 850–854 (2002).
27. Ohsumi, K., Koyanagi, A., Yamamoto, T. M., Gotoh, T. & Kishimoto, T. Emi1-mediated M-phase arrest in *Xenopus* eggs is distinct from cytostatic factor arrest. *Proc. Natl Acad. Sci. USA* **101**, 12531–12536 (2004).
28. Reimann, J. D., Gardner, B. E., Margottin-Goguet, F. & Jackson, P. K. Emi1 regulates the anaphase-promoting complex by a different mechanism than Mad2 proteins. *Genes Dev.* **15**, 3278–3285 (2001).
29. Moreno, S., Klar, A. & Nurse, P. Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**, 795–823 (1991).
30. Maundrell, K. *nmt1* of fission yeast. A highly transcribed gene completely repressed by thiamine. *J. Biol. Chem.* **265**, 10857–10864 (1990).

**Acknowledgements** We thank K. Tsumoto and M. Toya for technical assistance, Y. Watanabe and Y. Imai for helpful discussion, M. Yanagida for anti-Cut2p antibodies, and T. Matsumoto for the *slp1-362* strain. This work was supported by a Grant-in-Aid for Specially Promoted Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to M.Y. (myamamoto@ims.u-tokyo.ac.jp).

## Recruitment of *Drosophila* Polycomb group proteins to chromatin by DSP1

Jérôme Déjardin<sup>1</sup>, Aurélien Rappailles<sup>2</sup>, Olivier Cuvier<sup>1</sup>, Charlotte Grimaud<sup>1</sup>, Martine Decoville<sup>2</sup>, Daniel Locker<sup>2</sup> & Giacomo Cavalli<sup>1</sup>

<sup>1</sup>Institute of Human Genetics, CNRS, 141 rue de la Cardonille, F-34396 Montpellier Cedex 5, France

<sup>2</sup>Centre de Biophysique Moléculaire, CNRS/Université d'Orléans, rue C. Sadron, F-45071 Orleans Cedex 2, France

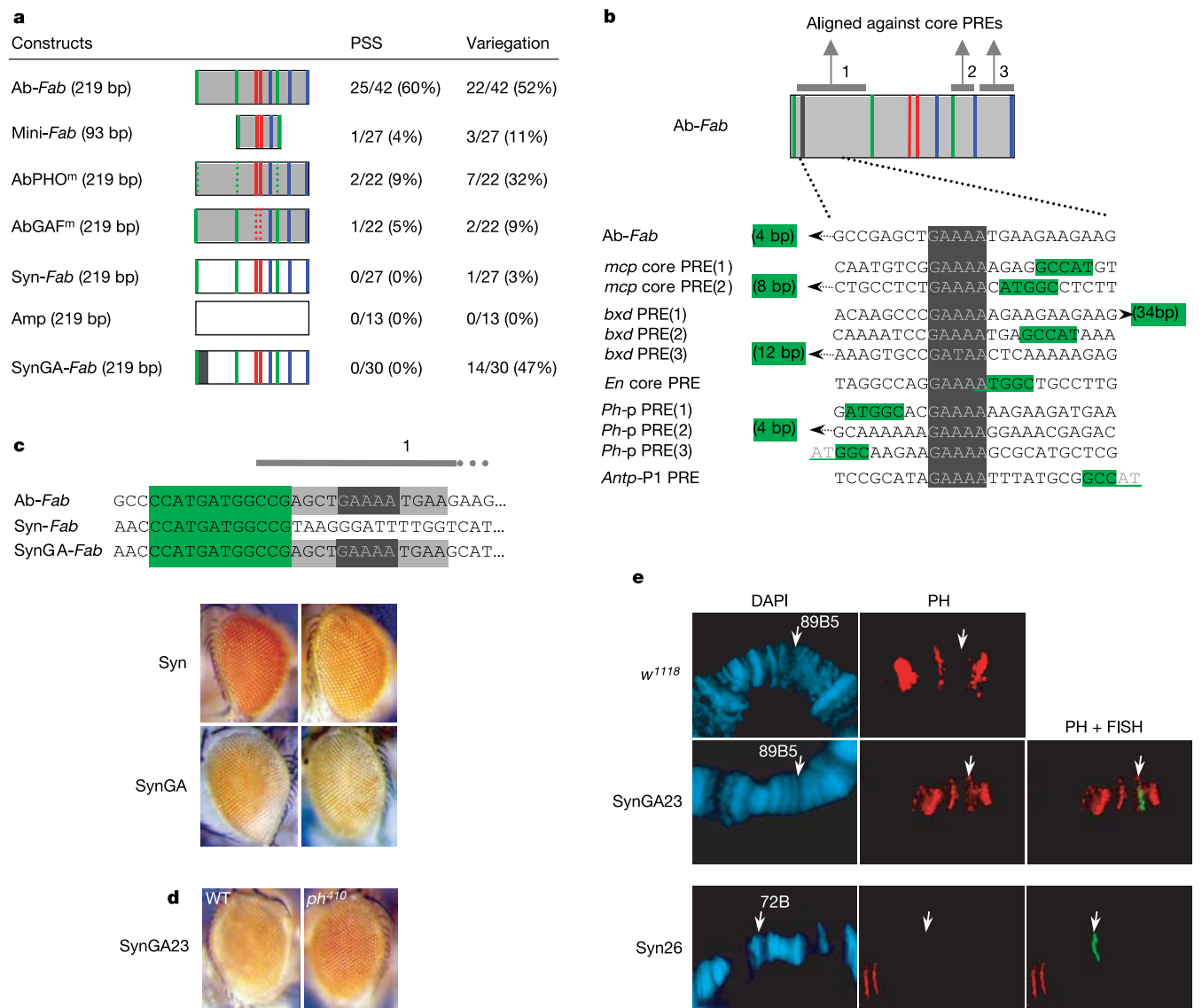
Polycomb and trithorax group (PcG and trxG) proteins maintain silent and active transcriptional states, respectively, throughout development<sup>1</sup>. In *Drosophila*, PcG and trxG proteins associate with DNA regions named Polycomb and trithorax response elements (PRE and TRE), but the mechanisms of recruitment are unknown. We previously characterized a minimal element from the regulatory region of the *Abdominal-B* gene, termed *Ab-Fab*. *Ab-Fab* contains a PRE and a TRE and is able to maintain repressed or active chromatin states during development<sup>2</sup>. Here we show that the Dorsal switch protein 1 (DSP1), a *Drosophila* HMGB2 homologue, binds to a sequence present within *Ab-Fab* and in other characterized PREs. Addition of this motif to an artificial sequence containing Pleiohomeotic and GAGA factor consensus sites is sufficient for PcG protein recruitment *in vivo*. Mutations that abolish DSP1 binding to *Ab-Fab* and to a PRE from the *engrailed* locus lead to loss of PcG protein binding, loss of silencing, and switching of these PREs into constitutive TREs. The binding of DSP1 to PREs is therefore important for the recruitment of PcG proteins.

Two classes of PcG protein complexes are recruited at PREs: the PRC2 complex contains the histone methyltransferase Enhancer of

Zeste (E(Z)), whereas the PRC1 complex contains Polycomb (PC), Polyhomeotic (PH), Posterior sex combs and dRing. *Drosophila* PREs have been characterized in transgenic assays<sup>3–9</sup>. They induce a variegated silencing of the mini-*white* reporter gene. Variegation is often accompanied by a phenomenon called ‘pairing-sensitive silencing’ (PSS), in which the expression of the mini-*white* reporter gene is lower in homozygous individuals than in heterozygotes (that is, the eye colour is lighter in the homozygotes). Variegation and PSS are reduced or suppressed in mutant background for PcG genes<sup>10</sup>. Sequence analysis has revealed three consensus binding sites for PcG or trxG proteins. One is bound by the Pleiohomeotic (PHO) and PHO-like proteins, the only members of the PcG known to bind DNA in a sequence-specific manner. PHO induces the recruitment of E(Z), which in turn might recruit PRC1 via the methylation of Lys 27 and Lys 9 in histone H3, followed by

recognition of these marks by the PC chromodomain<sup>11</sup>. A second consensus site is bound by the GAGA factor (GAF). GAF is required for PSS<sup>7</sup>, but its mechanism of action is unclear. The third motif is the consensus site for Zeste, a trxG protein<sup>12</sup>. PHO and GAF motifs seem to be necessary but not sufficient for PcG recruitment at PREs<sup>9,13–15</sup>, indicating that other factors might be involved.

We have recently characterized a core PRE/TRE of 219 base pairs (bp), named Ab-*Fab* (ref. 2). Ab-*Fab* is able to induce PcG-dependent PSS and PcG protein recruitment to transgenes in polytene chromosome staining assays. Ab-*Fab* contains binding sites for PHO, GAF and Zeste. Whereas mutation of the Zeste motifs selectively affects the trxG response<sup>2</sup>, we found that mutations in the PHO or GAF consensus sites disrupt PSS (AbPHO<sup>m</sup> or AbGAF<sup>m</sup>; see Fig. 1a). Mutation of PHO sites resulted in the complete loss of



**Figure 1** The G(A) motif contributes to the recruitment of PcG proteins. **a**, Fragments derived from Ab-*Fab*, and their ability to repress mini-*white* in a pairing-dependent fashion (PSS) and to display eye colour variegation. Green, red and blue bars represent binding sites for PHO, GAF and Zeste, respectively. **b**, Sequence alignments identify the G(A) motif in *Drosophila* PREs from Ab-*Fab* region 1. Arrows near the sequences indicate the distance in base pairs separating the edge of the alignment to the closest PHO

consensus site (highlighted in green). **c**, 5' sequence of Ab-*Fab*, Syn and SynGA-*Fab* constructs (region 1), and typical examples of mini-*white* expression produced by Syn or SynGA-*Fab* fragments. **d**, Effect of *ph*<sup>410</sup> background on SynGA-*Fab*-mediated silencing of mini-*white*. **e**, Immuno-FISH analysis with an anti-PH antibody on polytene chromosomes from transgenic lines of Syn26 and SynGA23 insertions. DAPI, 4,6-diamidino-2-phenylindole.



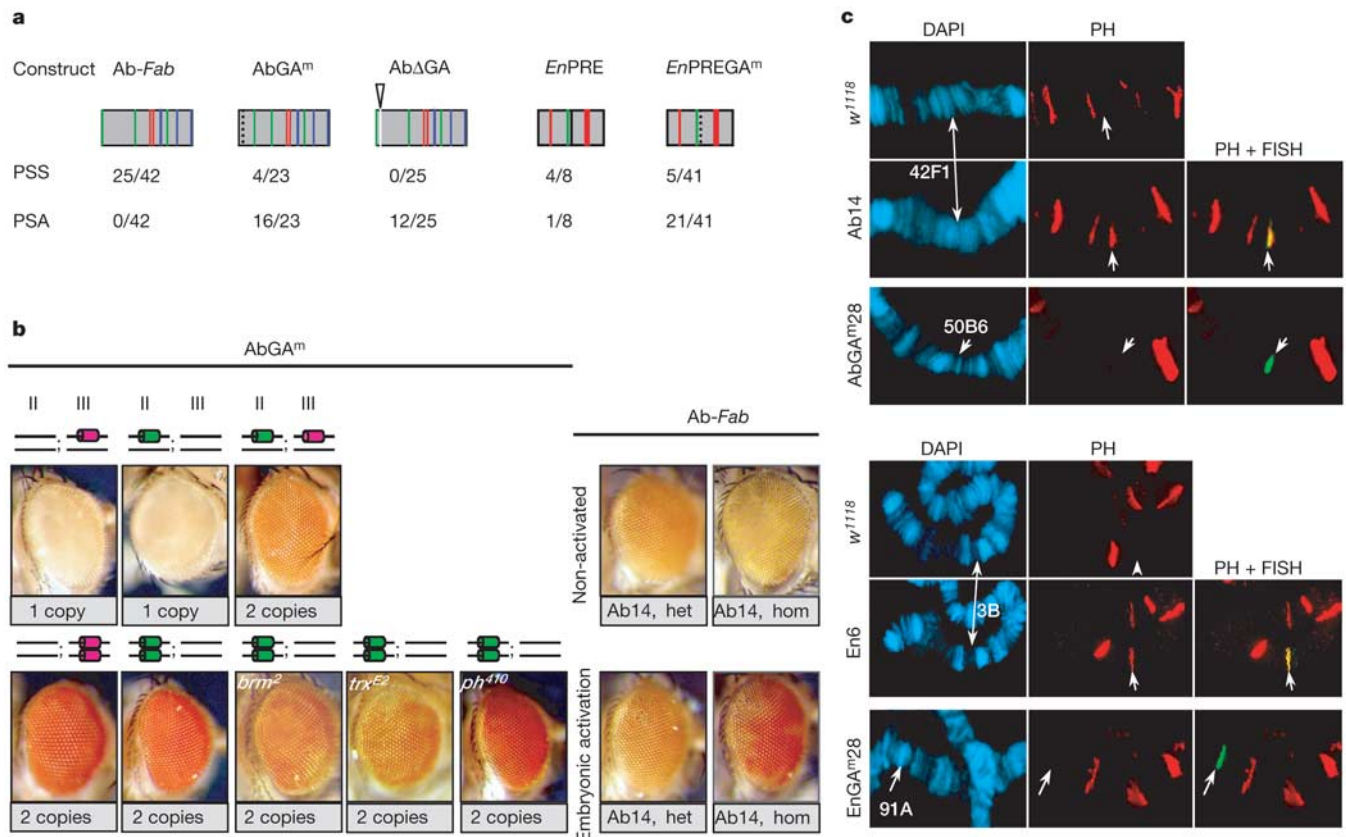
PH association with the transgene on polytene chromosomes (see Supplementary Fig. 1), which is consistent with a role for PHO in PcG recruitment<sup>16</sup>. Mutation of GAF sites significantly lowered, but did not disrupt, PH recruitment (see Supplementary Fig. 1), indicating that GAF binding sites might be partly dispensable for PcG recruitment.

To determine whether these consensus sites are sufficient to define the minimal *Ab-Fab* PRE, we designed an artificial sequence named *Syn-Fab*. *Syn-Fab* contains all PHO, GAF and Zeste consensus sites in the same order, orientation and spacing as in *Ab-Fab*, but with the intervening sequences replaced by bacterial sequences devoid of these sites. As a control, we used 219bp of the same bacterial sequences (Fig. 1a, Amp construct). Neither of these elements was able to induce PSS (Fig. 1a) or eye variegation in transgenic lines. Moreover, *Syn-Fab* was unable to induce ectopic binding of the PH protein to polytene chromosomes (Fig. 1e). Together with previously published evidence<sup>15–17</sup>, these results strongly indicate that binding sites for PHO, GAF and Zeste are not sufficient to define a PRE.

The *Mini-Fab* construct (93bp) was also unable to induce variegation and PSS, or to recruit PcG proteins (Fig. 1a and ref. 2). This narrowed down the additional PcG recruiter DNA motifs to three blocks of *Ab-Fab* DNA that are located outside *Mini-Fab* and are not PHO, GAF or Zeste consensus motifs (regions 1, 2 and 3; see Fig. 1b). Sequence alignment between these blocks and the known core *Drosophila* PREs allowed us to identify the

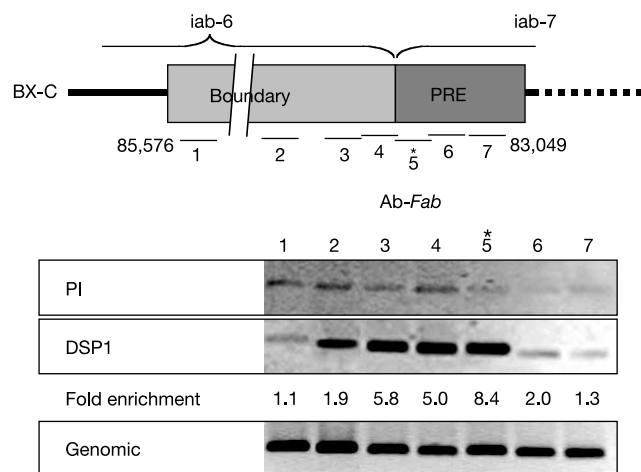
5'-GAAAA-3' candidate motif (hereafter called the G(A) motif), which is present in all molecularly characterized PREs at positions close to or overlapping PHO sites (Fig. 1b). Strikingly, the addition of a 14-bp sequence containing the G(A) site to *Syn-Fab* induced variegation of *mini-white* in 47% of the *SynGA-Fab* lines (Fig. 1a, c). Introducing the *SynGA23* transgenic flies in a *ph* mutant background (*ph*<sup>410</sup>) induced derepression of *mini-white*, indicating that the variegation is PcG-dependent (Fig. 1d). Moreover, we observed recruitment of PH protein to the insertion site of the transgene in polytene chromosomes of two different *SynGA-Fab* insertions (Fig. 1e).

To analyse the mechanism by which the G(A) motif contributes to the recruitment of PcG proteins, we analysed the binding of the two known PcG recruiters, PHO and GAF. Surprisingly, although containing clusters of PHO and GAF binding sites, *Syn-Fab* does not recruit these proteins *in vivo* (see Supplementary Figs 2 and 3). Whereas the presence of the G(A) motif in *SynGA-Fab* did not induce recruitment of GAF (Supplementary Fig. 3), it induced weak but significant binding of PHO (Supplementary Fig. 2). Consistent with a role of PHO in the recruitment of E(Z) was our observation that PHO binding was paralleled by recruitment of E(Z) to *SynGA-Fab* (see Supplementary Fig. 4). Thus, the addition of the G(A) motif is sufficient to recruit the PcG proteins PHO, E(Z) and PH *in vivo* in the context of a heterologous DNA sequence, although the absence of PSS and the weaker binding observed in the synthetic PRE than in the natural *Ab-Fab* sequence indicate that other *Ab-Fab*



**Figure 2** The G(A) motif is required for PcG-mediated silencing and for PH recruitment in natural PREs. **a**, Effect of wild-type or G(A) mutated *Ab-Fab* and *EnPRE* on PSS or PSA of *mini-white* expression. The number of lines showing each effect is indicated together with the total number of lines. **b**, Analysis of the *trxG*-dependent PSA. Left: the PSA effect on the two representative *AbGA*<sup>m</sup>14 and *AbGA*<sup>m</sup>28 lines. Eye pigment levels of homozygous individuals are, respectively, 13-fold and 20-fold those of their heterozygous siblings. PSA

observed in *AbGA*<sup>m</sup>28 is lost in *trx*<sup>E2</sup> or *brm*<sup>m2</sup> mutant backgrounds. Right: when heterozygous, the *Ab-Fab* insertion does not maintain the memory of active chromatin states after a pulse of embryonic GAL4 activation. **c**, Immuno-FISH with an anti-PH antibody in lines carrying the *Ab-Fab*, *AbGA*<sup>m</sup>, *EnPRE* and *EnPREGA*<sup>m</sup> constructs. DAPI, 4,6-diamidino-2-phenylindole.



**Figure 3** Binding of DSP1 in the *Fab-7* region. DSP1 ChIP performed on embryos 0–14 h old. The association of DSP1 to seven fragments of about 220 bp covering the *Fab-7* element is shown. PI indicates the preimmune serum control immunoprecipitation. Fold enrichment indicates the ratio between DSP1 and PI signals. The panel labelled 'genomic' shows the amplification of genomic DNA with the same primers, done as a control for the efficiency of PCR amplification. Note that, although the G(A)-containing *Ab-Fab* region is strongly bound by DSP1, a neighbouring fragment containing four G(A) motifs is weakly enriched (PCR fragment 6). This indicates that the presence of a G(A) motif is not the sole determinant of binding. The DNA sequence context might influence the strength of DSP1 binding *in vivo*, or binding might require additional DNA sequence determinants. The asterisk indicates the *Ab-Fab* fragment. *iab-6* and *iab-7* indicate two genomic regions regulating the parasegment-specific expression of the homeotic gene *Abd-B*. The endogenous *Fab-7* element is located at the border between them.

DNA sequences might contribute to robust tethering of PcG proteins and to the pairing sensitivity.

To evaluate the importance of the G(A) motif, we mutated or deleted it in two natural PREs: *Ab-Fab* and the core PRE from the *engrailed* gene (*EnPRE*)<sup>15</sup> (constructs *AbGA<sup>m</sup>* or *AbΔGA* and *EnPREGA<sup>m</sup>*; Fig. 2a). These mutations disrupted PSS, and they severely reduced the recruitment of PH, PHO and E(Z) proteins to the transgenes (see Fig. 2 and Supplementary Figs 2 and 4). This shows that the G(A) motif is necessary for PcG recruitment at *Drosophila* PREs. In contrast, GAF was still able to bind the *AbGA<sup>m</sup>-Fab* transgene (see Supplementary Fig. 3), indicating that the GAF pathway of PcG protein recruitment can be uncoupled from the PHO pathway.

In addition to loss of silencing, an unexpected phenotype was observed in about half of the lines carrying G(A) mutated transgenes: *mini-white* expression was abnormally increased when the transgenes were in the homozygous state. We call this phenotype 'pairing-sensitive activation' (PSA), to indicate that *mini-white* expression is synergistically activated by transgene homozygosity, resulting in eye pigment levels much higher than double the heterozygous levels (Fig. 2b). Crossing flies from two independent *AbGA<sup>m</sup>* PSA lines results in *trans* heterozygous progeny (carrying one copy of the transgene at each of two distinct insertion sites). In this case, the two transgenes express additively instead of synergistically (Fig. 2b), indicating that, as in PSS, PSA depends on homologous pairing of the transgenes and not simply on the total transgene copy number. PSA also depends on *trxG* proteins, because it was affected in an *AbGA<sup>m</sup>* line tested in a mutant background for the two *trxG* genes *trithorax* (*trx*) or *brahma* (*brm*). Conversely, mutation in *ph* (*ph<sup>410</sup>*) did not have any effect (Fig. 2b). *Ab-Fab* has previously

been shown to mediate the *trxG*-dependent maintenance of active chromatin states after an embryonic pulse of transcriptional activation<sup>2</sup>. However, we found that this maintenance can be observed only when *Ab-Fab* is in a homozygous state (Fig. 2b), indicating that somatic pairing of homologous chromosomes, a phenomenon known for its influence on homeotic gene expression in *Drosophila*<sup>18</sup>, might affect *trxG*-dependent regulation. These data indicate that mutation of the G(A) motif mimics an embryonic pulse of transcriptional activation, switching a PRE into a TRE.

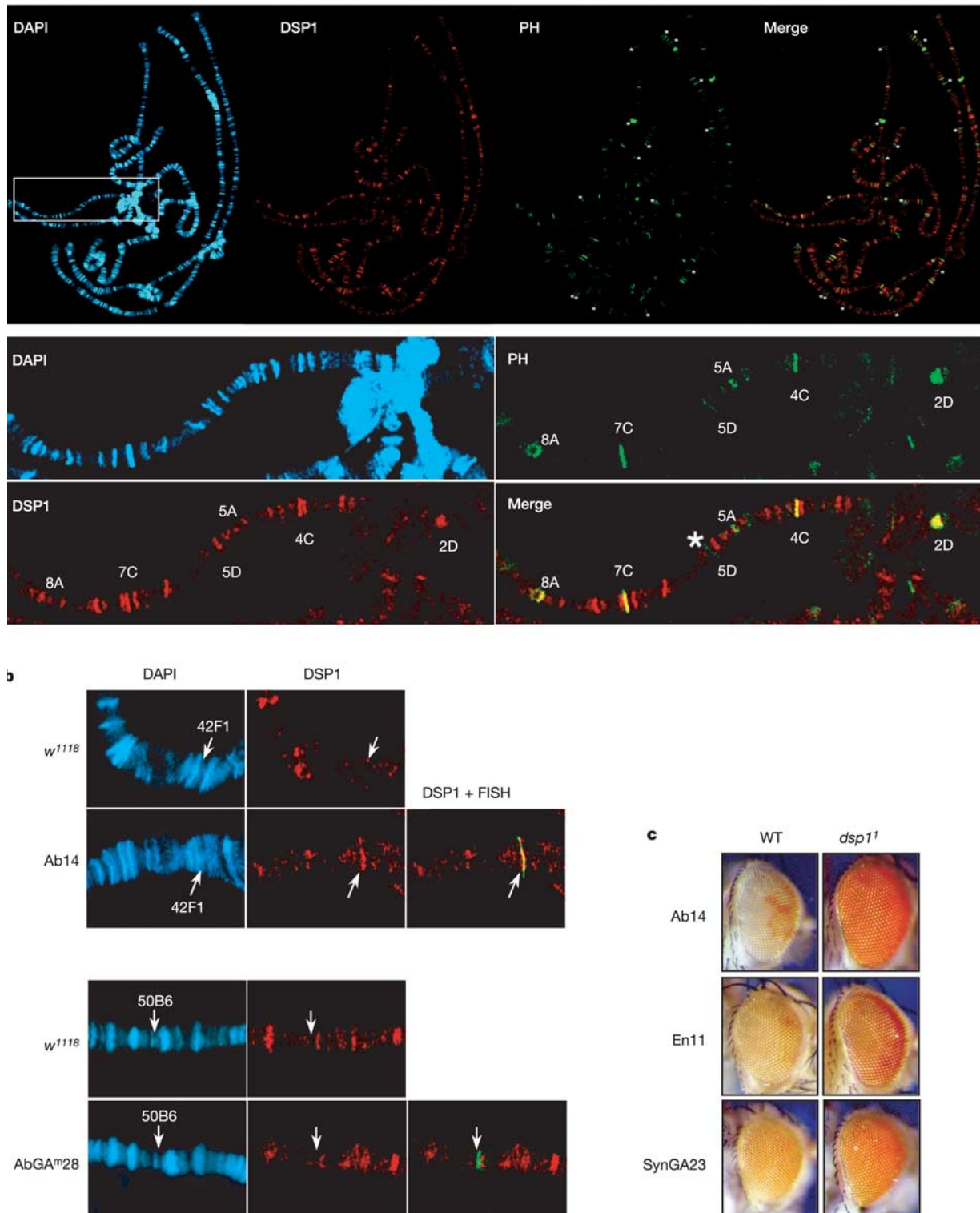
The G(A) motif is specifically bound by a nuclear activity in electrophoretic mobility-shift assays (EMSAs) with an 18-bp probe from *Ab-Fab* (see Supplementary Fig. 5). Among candidate proteins that might recognize this motif, we analysed DSP1, which is a *Drosophila* homologue of human HMGB2. This HMG protein has been shown to act as a transcriptional corepressor<sup>19,20</sup> and to be capable of recognition of DNA motifs resembling the G(A) site<sup>20</sup>. Moreover, DSP1 is required for the generation of a DNase I-hypersensitive region at the PRE-containing MCP element of the *BX-C<sup>21</sup>*, a *Drosophila* Hox gene cluster. Last, a *dsp1* mutant displays PcG and *trxG* homeotic phenotypes<sup>21</sup>. Despite the fact that HMGB1 and HMGB2 show weak sequence-specificity of DNA binding *in vitro*<sup>22</sup>, EMSAs indicate that recombinant DSP1 induces the same shift of the G(A) probe as that obtained with nuclear extract, whereas it binds poorly to the mutated probe (see Supplementary Fig. 5). Moreover, chromatin immunoprecipitation (ChIP) experiments with embryonic chromatin showed that DSP1 binds the *Fab-7* region *in vivo*. Within *Fab-7*, binding was greatest at the *Ab-Fab* region (polymerase chain reaction (PCR) fragment 5; see Fig. 3), which is consistent with the previously reported binding profile of the Polycomb protein<sup>23</sup>. In polytene chromosomes, DSP1 binds specifically to about 200 distinct sites (Fig. 4a). In double-staining experiments, DSP1 localizes together with the PH protein at most of the PH target sites (Fig. 4a), although the absence of PH from a fraction of the DSP1 sites indicates that DSP1 has other functions in addition to PcG protein recruitment.

DSP1 was recruited to the *Ab-Fab* transgene *in vivo*, and mutation of the G(A) motif in the *AbGA<sup>m</sup>* transgene abolished DSP1 recruitment (Fig. 4b). The *EnPRE* transgene also induced a G(A)-motif-dependent DSP1 recruitment *in vivo*, although to weaker levels than *Ab-Fab*. Together with the *in vitro* data, these results indicate that DSP1 binding to PREs induces the recruitment of PcG proteins and PcG-mediated silencing. Consistent with this hypothesis, crossing of *Ab-Fab*, *SynGA-Fab* as well as *EnPRE* transgenic flies into a *dsp1* loss-of-function mutant background (*dsp1<sup>1</sup>*) induced strong derepression of *mini-white*, to levels similar to those of PSA lines carrying transgenes with mutated G(A) motifs (see Fig. 4c). Derepression was particularly strong in males homozygous for the transgene, whereas it was weaker in heterozygous individuals (see Supplementary Fig. 6). This paralleled a strong decrease in PH recruitment, as seen by polytene chromosome staining. Intriguingly, these effects were less pronounced in females (see Supplementary Fig. 7), indicating that other factors might partly complement for DSP1 in female flies.

The role of DSP1 might be to help establish a chromatin architecture that facilitates the binding of PHO at its cognate consensus sequences in PREs. Once bound, PHO might then recruit the PRC2 complex, as well as PRC1 components, by direct protein–protein contacts and induction of histone marks<sup>11,24</sup>. The identification of a G(A) motif in a fraction of putative PRE sequences identified in a bioinformatic survey of the *Drosophila* genome<sup>12</sup> indicates that a DSP1-dependent mechanism might be widely used to recruit PcG proteins at their genomic targets. In contrast, GAF recruitment is independent of DSP1 and might require additional DNA sequences present within the core PRE. The findings that

SynGA-*Fab* is unable to recruit GAF and that it does not induce pairing effects indicate that GAF might contribute to pairing-dependent functions of *Drosophila* PREs. In humans, DNA repetitive sequences named 'D4Z4 repeats' repress neighbouring genes by

the recruitment of a repressive complex containing Yin-Yang-1 (the human homologue of PHO), nucleolin and HMGB2 proteins<sup>25</sup>. The link between HMGB2 proteins and the PcG machinery might therefore be evolutionarily conserved. □



**Figure 4** DSP1 participates in PcG-mediated repression. **a**, Co-localization between PH and DSP1 on polytene chromosomes from *w<sup>1118</sup>* larvae. Asterisks indicate PH sites not bound by DSP1. The bottom panel shows a magnification of the X chromosome, in which all PH sites (except 5D) are also bound by DSP1. DAPI, 4,6-diamidino-2-phenylindole.

**b**, Ab-*Fab*, but not AbGA<sup>m</sup>, recruits DSP1 *in vivo*, as shown by DSP1 immuno-FISH on Ab14 or AbGA<sup>m</sup>28 larvae. **c**, Derepression of mini-*white* in the Ab14, En11 and SynGA23 transgenic lines in a *dsp1<sup>1</sup>* mutant background. WT, wild type.



## Methods

### DNA cloning, synthetic *Fab* constructs and fly work

Ab-*Fab*, Mini-*Fab*, AbPHO<sup>m</sup>, AbGAF<sup>m</sup>, AbGA<sup>m</sup>, AbΔGA, *EnPRE* and *EnPREGA*<sup>m</sup> fragments were obtained by PCR with specific primers. *Syn-Fab* was constructed by primer extension by using five overlapping oligonucleotides. An additional round of PCR was performed on the resulting template. *SynGA-Fab* was obtained after PCR mutagenesis on the *Syn-Fab* template. The AbGA<sup>m</sup> mutant bears the GACGC mutation in place of GAAAA, whereas the *EnPREGA*<sup>m</sup> mutant bears the GCGCA motif so as to leave the PHO site intact. In AbΔGA, the GAAAA motif was deleted from Ab-*Fab*. The sequences of all primers are available from the authors on request. All inserts are cloned into the pUZ vector for transgenesis and were checked by sequencing. Insertion orientation and insert copy number were also verified by PCR. Details of crosses used to obtain transgenic lines into *trx*<sup>E2</sup>, *brm*<sup>2</sup>, *dsp1*<sup>1</sup> and *ph*<sup>410</sup> mutant backgrounds are available from the authors on request. Eye pictures were obtained using a Nikon Coolpix 990 charge-coupled device camera mounted on a MZFLIII binocular microscope (Leica). For eye colour comparison, all individuals to be compared were of the same age and were imaged in the same frame.

### Immunostaining and immuno-FISH on polytene chromosomes

Both chromosome staining procedures were performed as described previously<sup>2</sup>. Antibodies used in immuno-FISH (fluorescence *in situ* hybridization) were described previously<sup>2,26</sup>. For DSP1/PH double immunostaining, a goat anti-PH antibody was developed with the same protocol as that for the rabbit anti-PH, and recognizes the same bands in polytene chromosomes.

### Nuclear extract preparation, DSP1 purification and EMSAs

Nuclear extracts from 0–12-h embryos were prepared as described previously<sup>27</sup>. Nuclear extracts obtained from *Drosophila* SL2 cells were prepared with the Active-Motif Nuclear Extract Kit, in accordance with the manufacturer's recommendations. Recombinant DSP1 protein was obtained as described previously<sup>26</sup>. Probe sequences used in the bandshift assays were as follows: WT (wild type), 5'-CGAGCTGAAATGAAGAA-3'; and Mut, 5'-CGAGCTGACGCTGAAGAA-3'. Double-stranded probes were labelled with [ $\gamma$ -<sup>32</sup>P]ATP and the T4 polynucleotide kinase (New England Biolabs) in accordance with the manufacturer's instructions. The mutation used is the same as that used *in vivo* into AbGA<sup>m</sup>. Binding reactions were performed for 20 min at 4 °C with 5  $\mu$ g nuclear extract and 25 fmol labelled probes (approx. 10<sup>4</sup> c.p.m. per reaction), in 15  $\mu$ l binding buffer (25 mM HEPES-KOH pH 7.6, 1  $\mu$ g poly(dI-dC), 10  $\mu$ M zinc acetate, 3 mM MgCl<sub>2</sub>, 30 mM KCl, 2.5 mM dithiothreitol, 0.01% Tween 20, 5% glycerol). DNA-protein complexes were resolved on 6% polyacrylamide-Tris/borate/EDTA gels at 4 °C and 10 V cm<sup>-1</sup>.

### ChIP experiments

Cross-linked embryonic chromatin was prepared as described previously<sup>23</sup> and immunoprecipitations were performed by following the protocol in ref. 28. Chromatin (50  $\mu$ g) was immunoprecipitated with either 1.5  $\mu$ g purified anti-DSP1 antibodies or with preimmune serum. The PCR fragment coordinates correspond to the published *BX-C* sequence<sup>29</sup>. Quantification was performed on the products from two independent ChIP experiments, and enrichments are expressed as the ratio between the anti-DSP1 and the preimmune serum signals. The sequence of PCR primers is available from the authors upon request.

Received 3 November 2004; accepted 25 January 2005; doi:10.1038/nature03386.

- Francis, N. J. & Kingston, R. E. Mechanisms of transcriptional memory. *Nature Rev. Mol. Cell Biol.* **2**, 409–421 (2001).
- Déjardin, J. & Cavalli, G. Chromatin inheritance upon Zeste-mediated Brahma recruitment at a minimal cellular memory module. *EMBO J.* **23**, 857–868 (2004).
- Fauvarque, M.-O. & Dura, J.-M. *polyhomeotic* regulatory sequences induce developmental regulator-dependent variegation and targeted P-element insertions in *Drosophila*. *Genes Dev.* **7**, 1508–1520 (1993).
- Kassis, J. A. Unusual properties of regulatory DNA from the *Drosophila engrailed* gene: three 'pairing-sensitive' sites within a 1.6-kb region. *Genetics* **136**, 1025–1038 (1994).
- Chan, C. S., Rastelli, L. & Pirrotta, V. A *Polycomb* response element in the *Ubx* gene that determines an epigenetically inherited state of repression. *EMBO J.* **13**, 2553–2564 (1994).
- Zink, D. & Paro, R. *Drosophila Polycomb*-group regulated chromatin inhibits the accessibility of a *trans*-activator to its target DNA. *EMBO J.* **14**, 5660–5671 (1995).
- Hagstrom, K., Muller, M. & Schedl, P. A *Polycomb* and GAGA dependent silencer adjoins the *Fab-7* boundary in the *Drosophila* bithorax complex. *Genetics* **146**, 1365–1380 (1997).
- Shimell, M. J., Peterson, A. J., Burr, J., Simon, J. A. & O'Connor, M. B. Functional analysis of repressor

- binding sites in the *iab-2* regulatory region of the abdominal-A homeotic gene. *Dev. Biol.* **218**, 38–52 (2000).
- Mishra, R. K. *et al.* The *iab-7* polycomb response element maps to a nucleosome-free region of chromatin and requires both GAGA and Pleiohomeotic for silencing activity. *Mol. Cell. Biol.* **21**, 1311–1318 (2001).
- Kassis, J. A. Pairing-sensitive silencing, polycomb group response elements, and transposon homing in *Drosophila*. *Adv. Genet.* **46**, 421–438 (2002).
- Wang, L. *et al.* Hierarchical recruitment of polycomb group silencing complexes. *Mol. Cell* **14**, 637–646 (2004).
- Ringrose, L., Rehmsmeier, M., Dura, J. M. & Paro, R. Genome wide prediction of Polycomb/Trithorax response elements in *Drosophila melanogaster*. *Dev. Cell* **5**, 759–771 (2003).
- Fritsch, C., Brown, J. L., Kassis, J. A. & Muller, J. The DNA-binding polycomb group protein pleiohomeotic mediates silencing of a *Drosophila* homeotic gene. *Development* **126**, 3905–3913 (1999).
- Busturia, A. *et al.* The MCP silencer of the *Drosophila* Abd-B gene requires both Pleiohomeotic and GAGA factor for the maintenance of repression. *Development* **128**, 2163–2173 (2001).
- Americo, J. *et al.* A complex array of DNA-binding proteins required for pairing-sensitive silencing by a Polycomb group response element from the *Drosophila* engrailed gene. *Genetics* **160**, 1561–1571 (2002).
- Brown, J. L., Mucci, D., Whiteley, M., Dirksen, M.-L. & Kassis, J. A. The *Drosophila* Polycomb group gene *pleiohomeotic* encodes a DNA binding protein with homology to the transcription factor YY1. *Mol. Cell* **1**, 1057–1064 (1998).
- Brown, J. L., Fritsch, C., Mueller, J. & Kassis, J. A. The *Drosophila* pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing. *Development* **130**, 285–294 (2003).
- Duncan, I. W. Transvection effects in *Drosophila*. *Annu. Rev. Genet.* **36**, 521–556 (2002).
- Lehming, N. *et al.* An HMG-like protein that can switch a transcriptional activator to a repressor. *Nature* **371**, 175–179 (1994).
- Brickman, J. M., Adam, M. & Ptashne, M. Interactions between an HMG-1 protein and members of the Rel family. *Proc. Natl Acad. Sci. USA* **96**, 10679–10683 (1999).
- Decoville, M., Giacomello, E., Leng, M. & Locker, D. DSP1, an HMG-like protein, is involved in the regulation of homeotic genes. *Genetics* **157**, 237–244 (2001).
- Thomas, J. O. HMG1 and 2: architectural DNA-binding proteins. *Biochem. Soc. Trans.* **29**, 395–401 (2001).
- Cavalli, G. & Paro, R. The *Drosophila Fab-7* chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell* **93**, 505–518 (1998).
- Mohd-Sarip, A., Venturini, F., Chalkley, G. E. & Verrijzer, C. P. Pleiohomeotic can link polycomb to DNA and mediate transcriptional repression. *Mol. Cell. Biol.* **22**, 7473–7483 (2002).
- Gabellini, D., Green, M. R. & Tupler, R. Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. *Cell* **110**, 339–348 (2002).
- Mosrin-Huaman, C., Canaple, L., Locker, D. & Decoville, M. DSP1 gene of *Drosophila melanogaster* encodes an HMG-domain protein that plays multiple roles in development. *Dev. Genet.* **23**, 324–334 (1998).
- Frank, A. *et al.* Polycomb and polyhomeotic are constituents of a multimeric protein complex in chromatin of *Drosophila melanogaster*. *EMBO J.* **11**, 2941–2950 (1992).
- Orlando, V., Strutt, H. & Paro, R. Analysis of chromatin structure by *in vivo* formaldehyde cross-linking. *Methods* **11**, 205–214 (1997).
- Martin, C. H. *et al.* Complete sequence of the bithorax complex of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **92**, 8398–8402 (1995).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This paper is dedicated to N. Riccò. We thank all laboratory members for input during the course of this work and for critical reading of the manuscript; F. Juge for developing the goat anti-PH antibody; P. Atger for artwork; R. Paro for the gift of rabbit anti-PH antibody; and V. Pirrotta for the gifts of anti-PHO and anti-E(Z). J.D. and A.R. were supported by fellowships from the 'Ministère de l'Enseignement Supérieur et de la Recherche' and by a grant from 'La ligue nationale contre le cancer'. O.C. is supported by an HFSP long-term fellowship. D.L. and M.D. are supported by grants from the CNRS. Research in the Cavalli laboratory is supported by grants from the CNRS, the HFSP, the 6th EU Framework Project, the 'Fondation pour la Recherche Médicale', the 'Fondation Schlumberger pour l'Éducation et la Recherche' and the 'Association pour la Recherche sur le Cancer'.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to G.C. (giacomo.cavalli@igh.cnrs.fr).

## Contacts

**Publisher:** Ben Crowe  
**Editor:** Paul Smaglik  
**Marketing Manager:** David Bowen

### European Head Office, London

The Macmillan Building  
4 Crinan Street  
London N1 9XW, UK  
Tel +44 (0) 20 7843 4961  
Fax +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

### Naturejobs Sales Director:

Nevin Bayouni (4978)

### European Sales Manager:

Andy Douglas (4975)

### UK/ RoW/ Ireland:

Nils Moeller (4953)  
Irene Vigila-Atton (4944)

### Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

**Natureevents:** Silke Opstrup (4994)

### France/ Switzerland:

Amelie Pequignot (4974)

### Advertising Production

**Manager:** Billie Franklin

To send materials use London  
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

### Naturejobs web development:

Tom Hancock

### Naturejobs online production:

Niamh Shields

### European Satellite Office

**Germany/ Austria/ Italy/  
The Netherlands/ Belgium:**

Patrick Phelan

Tel +49 89 54 90 57 11

Fax +49 89 54 90 57 20

e-mail: [p.phelan@nature.com](mailto:p.phelan@nature.com)

Reya Silao

Tel +44 (0) 20 7843 4970

e-mail: [r.silao@nature.com](mailto:r.silao@nature.com)

### US Head Office, New York

345 Park Avenue South,  
10th Floor, New York, NY 10010-1707  
Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

### US Sales Manager: Peter Bless

### Japan Head Office, Tokyo

MIG Ichigaya Building (5F),  
19-1 Harakutamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

### Asia-Pacific Sales Director:

Rinoko Asami

e-mail: [rasami@naturejpn.com](mailto:rasami@naturejpn.com)

# naturejobs

## Tale of two surveys

Two surveys unveiled at the National Postdoctoral Association (NPA) annual meeting in San Diego this month showed progress for postdocs on issues of salary and job satisfaction. But they also highlighted current problems, and indicated storm clouds on the horizon.

The first survey, conducted by Sigma Xi, in conjunction with the NPA, showed that 40% of the 7,600 postdocs questioned earned an average stipend of US\$38,000 — up from \$28,000 in 1995. Physicists populated the higher end of the pay scale and ecologists the lower, but the majority of them expressed positive job satisfaction. The bad news comes from the postdocs who received their PhDs outside the United States — on average, they worked more hours and published more papers, but were on the lower end of the pay scale. And there were concerns about visas, with 30% of foreign postdocs saying national-security issues affected their research and 57% saying that security regulations made it harder to re-enter the United States.

The other survey, by chemist Joseph Cerny at the University of California, Berkeley, looked at how nuclear chemists and nuclear physicists viewed their training five and ten years after receiving their doctorates. The good news is that almost all were glad they did a PhD. The bad news is that only a third ended up working in the kind of job they had anticipated. And the trouble on the horizon? It was the quality of mentoring that led people to the wrong job, many respondents said. Furthermore, says Cerny, the time to tenure in some disciplines, especially biochemistry, is getting longer, and prospects may be even worse for oversubscribed disciplines such as molecular biology.

Still, postdocs everywhere can rejoice that these surveys have become public — and hope that they effect changes for the better.

**Paul Smaglik**  
*Naturejobs editor*



## Contents

### SPECIAL REPORT

The plight of Europe's  
postdocs

p540

### CAREER VIEW

**Recruiters & Industry**  
Hard times for Swedish  
physician-researchers  
**Graduate Journal**  
Facing the reviewers  
**Movers**

Carole Moquin-Pathey

p542

### WWW.NATUREJOBS.COM

Career centre  
Information on the  
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

# Many European PhD students and junior researchers are getting a bad deal, with few rights and little or no supervision. But things are about to change, reports Quirin Schiermeier.

Sally Maindew's (not her real name) research career ended in tears and rage. In January, the 29-year-old Irish biochemist, working in Austria, told her supervisor that she was sick of the semi-legal, orally agreed working arrangements. She was pregnant and needed a proper working contract, including maternity benefits and social security. She will never forget his answer: "Please empty your office and hand me the keys."

On her way home, she thought of ending the pregnancy — a drastic idea. Instead she decided to fight her corner. But in the weeks that followed, she realized that without a written contract she had virtually no rights in Austria. Despite being a European Union (EU) citizen and working in an EU country, she was basically an outcast.

Half a million junior researchers, such as PhD students and postdoctoral workers, form the backbone of Europe's scientific pursuits. How many of them are employed under ill-defined conditions can only be guessed. Only a few young scientists may have experienced treatment as extreme as Maindew's, but many postgraduate researchers throughout the EU would agree that their supervision, education and legal status could be improved.

Associations that represent postgraduates are addressing these issues. Some organizations and scientific administrators would like more clearly defined conditions. Some are calling for standardization of training and conditions throughout Europe.

"PhD students are professionals, and they can expect to be treated with due respect," says Dagmar Meyer, an assistant professor of mathematics at the University of Göttingen in Germany and head of the Marie Curie Fellowship Association. "Unfortunately, they have too little say, and their needs and problems are not always taken seriously."

Recognition of a PhD, its value on the non-academic job market, and the structure of training undertaken vary from country to country.

In Germany and Austria, for example, a PhD is an essential requirement for senior positions in many non-academic professions. PhD holders proudly indicate their degree on letterheads, business cards and door-plates. This would be considered unusual in France, Spain or Britain.

"French companies don't have much interest in a PhD title," says Francis Vella, a French developmental biologist at the UK Medical Research Council's Clinical Sciences Centre in London. In France and Spain, PhD experience is not always valued, he says. "It can actually be a disadvantage on the job market, and so many applicants don't even mention it in their CVs."

In most EU countries, universities produce more PhDs than the academic system can absorb. As a result, an estimated two-thirds of PhD candidates need to look for a job in industry. "It would be better if PhDs were more generally recognized as a qualification for either career option," says Vella. Doctoral education should always include training in skills required outside the lab, he adds.

Such 'transferable' skills, for example project management and budgeting, have already become an integral part of many doctoral programmes. Several countries have also begun to establish formal requirements for doctoral training, to ensure that candidates get appropriate support and supervision, and finish their PhDs in a reasonable time.

The British Quality Assurance Agency for Higher Education, for example, has defined the skills and competencies required to be awarded a PhD.

In Britain and Ireland, contractual regulations outlining the duties and obligations of PhD students and their supervisors have decreased drop-out

rates and helped reduce the length of study to three years. In other countries, such as Poland, Romania and Russia, excessive state validation continues to plague many junior researchers.

No wonder that doctoral programmes and Marie Curie fellowships that allow scientists to work in Britain, Scandinavia and the Netherlands are attractive to people from central and eastern Europe, as well as to young researchers from Mediterranean countries. "Doctoral education in Spain is poorly structured, and there are no employment rights at all," says Toni Gabaldón, who recently moved from Valencia to the University of Nijmegen in the Netherlands to complete his PhD (see *Nature* **428**, 448–449; 2004).

"Here I have a proper contract and a mentor who could mediate if there were any problems with the supervision. I can also attend academic writing courses, and the university even gave me some money

**In Britain and Ireland, regulations outlining the duties and obligations of PhD students and supervisors have decreased drop-out rates and helped reduce the length of study.**



## IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

C IMAGES.COM/CORBIS

for my move and helped me find a place to live,” says Gabaldón.

Gabaldón coordinates the mobility workgroup of the European Council of Doctoral Candidates and Junior Researchers (EURODOC). At their annual meeting in Strasbourg in March, EURODOC delegates agreed on a charter for supervision and training for junior researchers that listed best practices for student-supervisor relations and methods for reviewing progress, and outlined complaint procedures for PhD candidates (see ‘The travel bug’, right).

If adopted, this charter should help researchers who have disagreements with their supervisors, or feel isolated in their departments, to get back on track, says Tim Brown, an assistant engineering professor at Aalborg University in Denmark, who edited the charter. Most of the EURODOC suggestions have already been taken up by the European Commission, which earlier this month published its own recommendations for a ‘career charter’ (see *Nature* 434, 127; 2005).

The ‘Bologna Process’ has even more ambitious goals. This is a joint initiative from European science ministers to standardize academic degrees across the continent. At a summit meeting this May in Bergen, ministers will discuss the creation of common structures for doctoral education in Europe.

But not everybody is convinced that this is necessary, or desirable. “Diversity is not such a bad thing,” says Brown. “An equal system of funding and supervision is hardly achievable anyway. So why not have a mosaic of models, as long as there are common objectives and minimum standards?”

A common PhD structure might be an advantage for some countries. In Austria and Germany, the relationship between candidates and supervisors is particularly close, which sometimes causes problems. Supervisors often expect teaching and other duties from PhD candidates, which can prolong their studies.

Nonetheless, in Germany, a survey carried out by Thesis, the German PhD students’ organization, revealed a great level of enthusiasm for their research and, despite problems, a high level of satisfaction with their supervisors.

This is no comfort for those researchers placed in an impossible situation. For young scientists such as Maindew, a new European charter for junior researchers is long overdue.

Quirin Schiermeier is *Nature’s* German correspondent.

### The travel bug

Mobility was more than a catchphrase at this month’s EURODOC assembly. Several of the delegates, including, ironically, at least one member of the mobility work group, had difficulty getting to Strasbourg.

Snežana Krstić, the member in question, has more travel problems than most people — and not just because of strikes by French railway and airport workers. In common with scientists from the Balkans, Ukraine, Belarus and Russia, she has to undergo a time-consuming procedure every time she travels to the European Union (EU). “It very much constrains our ability to participate in scientific meetings and communicate with other scientists,” she complains.

There is no spot left unstamped in Krstić’s passport. The young teaching assistant at the University of Belgrade in Serbia, who last year completed her PhD in chemical engineering, needs a visa to travel. The time it takes her to get the sought-after stamps varies depending on the country and the season. During the summer, for example, she says that people queue for hours outside the Greek embassy in Belgrade. And some embassy staff are not renowned for their politeness: “Sometimes I don’t even try to apply for a visa because I know what is waiting for me,” she says.

Like many young scientists from Serbia, she is eager to gain experience in other countries, but the cumbersome visa requirements make this difficult.

The European Commission has begun to address this issue. It has recommended that all EU countries introduce a fast-track procedure for issuing short-term visas for scientists travelling to meetings and conferences. The commission has also drafted a directive, to be issued in the summer, to ease the immigration of scientists who opt for longer-term work in the EU.

## GRADUATE JOURNAL

## Facing the reviewers

It is somewhat disheartening to reflect on my years of research and then look at the figures I've assembled for my first real research paper. I am proud of my near-finished work, but all those pages of failed experiments are hard to swallow, when I look at just seven pages of pretty pictures.

I can see the hurdles that remain, and most of them involve what I'm doing right now — staring at my laptop. Despite overwhelming mental distraction, I try to focus on writing this paper and imagining the difficult questions that those notorious reviewers might ask. I feel as if I'm 13 and being sent to the principal for smoking, except now it takes four weeks to find out my punishment. I am probably too close to this work to judge its merit: I know the intricacies of the experiments and that the story has some holes here and there.

During my graduate years, I have reviewed dozens of papers for my principal investigator and ripped apart data, suggested experiments or controls, and blatantly stated that this work is not worthy of the journal to which its been submitted. I am scared of getting a reviewer like myself! I'm hoping I am being too hard on myself because I know the work too well. In any case, I hope that the principal is in a good mood and I just get a slap on the wrist.

**Jason Underwood is a graduate student in microbiology at the University of California, Los Angeles.**

## RECRUITERS &amp; ACADEMIA

## Hard times for Swedish physician-researchers

**P**hysician-researchers form an important link between pre-clinical and clinical medicine. They often put research findings into practice and they play a crucial role in maintaining high standards in health care. They often teach and encourage medical students to do research. So it is important for them to strike a balance between medicine and research.

This isn't always easy, as a 2003 Swedish research council survey indicated. Many of the 425 medical students and 182 junior doctors cited bad pay and a heavy workload as reasons to abstain from research.

The situation is like that in North America: only one in five Swedish physicians with a PhD plans to become a senior lecturer. The number of researchers with MDs is declining, for several reasons.

Getting a PhD involves a huge effort, especially for MDs, who often have to work full-time in a clinical setting while doing their PhD. The average age for MDs at the Karolinska Institute to attain a PhD increased from 38.5 in 1991 to 40.8 in 1997. Sweden's healthcare system is constantly slashing costs and increasing patient numbers. MDs find it hard to allocate time for research, and the pace has made supervisors less keen to take on new recruits.

To become a specialist in Sweden, an MD has to complete a two-year internship and another five years of specialization. Their salary increases by US\$1,000 a month when they qualify, but that pay rise is often delayed for physician-researchers by the number of months they spent on research.

A PhD is no longer a prerequisite for top clinical positions at university and teaching hospitals. And

funding for medical research is declining.

To encourage more physician-scientists, economic incentives need to be created. Norway sets an example: junior doctors are allowed to include 12 months' research towards their specialization requirement. This would reduce the salary gap with colleagues who are not doing research. Healthcare providers should reward research and create an atmosphere where it is an integrated part of clinical work. This means creating positions for young physician-scientists and remedying the shortage of supervisors needed to keep Nordic medical research at a world-class level.

Finally, the state should increase funding for clinical research. If it doesn't, young physician-scientists may soon be another species menaced by extinction.

**Jonas F. Ludvigsson is a physician-scientist at Örebro University Hospital in Sweden.**

## MOVERS

Carole Moquin-Patthey, head of unit, European Medical Research Councils, Strasbourg, France



**T**he jewellery designed by Carole Moquin-Patthey embraces duality: a brooch may transform into a necklace, and small studs into long elegant earrings. Moquin-Patthey's own career is marked by a series of double lives — art and science, chemistry and biology, adaptability and obstinacy.

It has led her from work in a hospital pharmacy to heading the European

Medical Research Councils (EMRC), via plant biology, marine chemistry and membership of the French ethics committee. Along the way she learned Indonesian and, with her banker husband, raised three children. "Not a typical career path," she laughs.

She dropped in on the Scripps Institution of Oceanography during a holiday after graduating as a pharmacist. That impulsive move led to four months' research with William Fenical on soft corals, and to the realization that her future lay in life sciences. An encounter at a Gordon research conference sent her home to France to do a PhD.

During a few years in Australia and Indonesia, a mix of stubbornness and opportunism kept her going when the science job outlook looked bleak.

"I was translating drug-approval documents, but couldn't get a full-time job," she recalls. "I took care of my children, and it was a very enjoyable

period." A successful sideline in jewellery design didn't, however, distract her when she had a chance to return to science.

Her efforts eventually secured her a postdoc in the United States. Back in France after a decade, a 40-year-old mother-of-three without a job offer, she was undeterred by advice to stay at home. One of many unsolicited letters persuaded Boehringer Mannheim to create a new position for her. From there it was a short step to the centre of national research funding at the medical research foundation (FRM) and policy making at the national institute of health and medical research (Inserm).

Her EMRC post involves creating a new pathway for clinical research in Europe with a single entry point: part of the European public-private initiative for developing new therapeutic tools. It uses her experience in science and legal and regulatory areas — and with her easel at home, she still finds time for art. ■

## CV

**2001–2004:** Head of scientific strategies and partnering, Inserm, Paris, France.

**1998–2001:** Scientific director, FRM,

Paris, France.

**1996–1998:** Coordinator, clinical research and development, Boehringer Mannheim, France.

**1992–1995:** Scientist, Enzon, New Jersey, USA.

**1990–1992:** Postdoc, chemistry department, Columbia University, New York, USA.

**1986:** PhD in pharmaco-chemistry, School of Pharmacy.

**1984:** PharmD, option Industry, School of Pharmacy, Paris, France.

# Heartwired

Love is the drug.

**Joe Haldeman**

Margaret Stevenson walked up the two flights of stairs and came to a plain wooden door with the nameplate 'Relationships, Ltd'. She hesitated, then knocked. Someone buzzed her in.

She didn't know what to expect, but the simplicity surprised her: no receptionist, no outer office, no sign of a laboratory. Just a middle-aged man, conservative business suit, head fashionably shaved, sitting behind an uncluttered desk. He stood and offered his hand. "Mrs Stevenson? I'm Dr Damien."

She sat on the edge of the chair he offered.

"Our service is guaranteed," he said without preamble, "but it is neither inexpensive nor permanent."

"You wouldn't want it to be permanent," she said.

"No." He smiled. "Life would be pleasant, but neither of you would accomplish much." He reached into a drawer and pulled out a single sheet of paper and a pen. "Nevertheless, I must ask you to sign this waiver, which relieves our corporation of responsibility for anything you or he may do or say for the duration of the effect."

She picked up the waiver and scanned it. "When we talked on the phone, you said that there would be no physical danger and no lasting physical effect."

"That's part of the guarantee."

She put the paper down and picked up the pen, but hesitated. "How, exactly, does it work?"

He leaned back, lacing his fingers together over his abdomen, and looked directly at her. After a moment, he said: "The varieties of love are nearly infinite. Every person alive is theoretically able to love every other person alive, and in a variety of ways."

"Theoretically," she said.

"In our culture, love between a man and a woman normally goes through three stages: sexual attraction, romantic fascination and then long-term bonding. Each of them is mediated by a distinct condition of brain chemistry.

"A person may have all three at once, with only one being dominant at any given time. Thus a man might be in love with his wife, and at the same time be infatuated with his mistress, and yet be instantly attracted to any stranger with appropriate physical characteristics."

"That's exactly..."

He held up a hand. "I don't need to know any more than you've told me. You've been married 25 years, you have an anniversary coming up ... and you want it to be romantic."

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

"Yes." She didn't smile. "I know he's capable of romance."

"As are we all." He leaned forward and took two vials from the drawer, a blue one and a pink one. He looked at the blue one. "This is Formula One. It induces the first condition, sort of a Viagra for the mind."

She closed her eyes and shook her head, almost a shudder. "No. I want the second one."

"Formula Two." He slid the pink vial towards her. "You each take approximately half of this, while in each other's company, and for several days you will be in a state of mutual infatuation. You'll be like kids again."

She did smile at that. "Whether he knows he's taken it or not?"

"That's right. No placebo effect."

"And there is no Formula Three?"

"No. That takes time, and understanding, and a measure of luck." He shook his head ruefully and put the blue vial away. "But I think you have that already."

"We do. The old-married-couple kind."

"Now, the most effective way of administering the drug is through food or drink. You can put it in a favourite dish, one you're sure he'll finish, but only after it's been cooked. Above a hundred degrees centigrade, the compound will decompose."

"I don't often cook. Could it be a bottle of wine?"

"If you each drink half, yes."

"I can force myself." She took up the pen and signed the waiver, then opened her clutch purse and counted out ten £100 notes. "Half now, you said, and half upon satisfaction?"

"That's correct." He stood and offered his

hand again. "Good luck, Mrs Stevenson."

The reader may now imagine any one of nine permutations for this story's end.

In the one the author prefers, they go to a romantic French restaurant, the lights low, the food wonderful, a bottle of good Bordeaux between them.

She excuses herself to go to the ladies' loo, the vial palmed, and drops her purse. When he leans over to pick it up, she empties the vial into the bottle of wine.

When she returns, she is careful to consume half of the remaining wine, which is not difficult. They are both in an expansive, loving mood, comrades these 25 years.

As they finish the bottle, she feels the emotion building in her, doubling and redoubling. She can see the effect on him, as well: his eyes wide and dilated, his face flushed. He loosens his tie as she pats perspiration from her forehead.

It's all but unbearable! She has to confess, so that he will know there's nothing physically wrong with him. She takes the empty pink vial from her purse and opens her mouth to explain...

He opens his hand and the empty blue vial drops to the table. He grabs the tablecloth...

They are released on their own recognition once the magistrate understands the situation.

But they'll never be served in that restaurant again. ■

*Joe Haldeman has been a science fiction writer since 1969. He teaches writing every fall semester at the Massachusetts Institute of Technology. His novels and short stories have won five Hugo and four Nebula Awards.*